TASK: PA29
CDRL: AC01
2 August 1996

INFORMAL TECHNICAL REPORT
For
SOFTWARE TECHNOLOGY FOR ADAPTABLE, RELIABLE SYSTEMS
(STARS)

Canvas Knowledge Acquisition Guidebook
Version 1.0

STARS-PA29-AC01/001/00

2 August 1996

CONTRACT NO. F19628-93-C-0130

Prepared for:
Electronic Systems Center
Air Force Systems Command, USAF
Hanscom, AFB, MA 01731-2816

Prepared by:
Lockheed Martin Tactical Defense Systems
9255 Wellington Road
Manassas, VA 20110-4121

Distribution Statement "A"
per DoD Directive 5230.24
Authorized for public release; Distribution unlimited.

TASK: PA29
CDRL: AC01
2 August 1996

Data Reference: STARS-PA29-AC01/001/00
INFORMAL TECHNICAL REPORT
Canvas Knowledge Acquisition Guidebook
Version 1.0

This document, developed under the Software Technology for Adaptable, Reliable Systems (STARS) program, is approved for release under Distribution "A" of the Scientific and Technical Information Program Classification Scheme (DoD Directive 5230.24) unless otherwise indicated. Sponsored by the U.S. Defense Advanced Research Projects Agency (DARPA) under contract F19628-93-C-0130, the STARS program is supported by the military services, SEI, and MITRE, with the U.S. Air Force as the executive contracting agent. The information identified herein is subject to change. For further information, contact the authors at the following mailer address:
delivery@tds-gn.lmco.com.

TASK: PA29
CDRL: AC01
2 August 1996

Data Reference: STARS-PA29-AC01/001/00
INFORMAL TECHNICAL REPORT
Canvas Knowledge Acquisition Guidebook
Version 1.0

Abstract

This guidebook describes the Canvas approach to systematic knowledge acquisition. Canvas synthesizes elements of two distinct methods: Scenario-Based Engineering Process (SEP) and Organization Domain Modeling (ODM). SEP provides knowledge acquisition methods and representation techniques. ODM provides a conceptual framework for data acquisition planning for the purposes of domain engineering. The guidebook incorporates extensive lessons learned from project experience in managing a large-scale knowledge acquisition effort in coordination with advanced technology development in the health-care domain.

TASK: PA29
CDRL: AC01
2 August 1996

Data Reference: STARS-PA29-AC01/001/00
INFORMAL TECHNICAL REPORT
Canvas Knowledge Acquisition Guidebook
Version 1.0

Principal Author(s):

Mark Simos, Organon Motives, Inc.	Date
Dean Allemang, Organon Motives, Inc.	Date
Charles Hammons, ScenPro, Inc.	Date
Lisa Mantock, ScenPro, Inc.	Date
Carol Klingler, Lockheed Martin Tactical Defense Systems	Date
Larry Levine, Organon Motives, Inc.	Date
Dick Creps, Lockheed Martin Tactical Defense Systems	Date

Approvals:

Teri F. Payton, Lockheed Martin Tactical Defense Systems Date
STARS Program Manager

(Signatures on File)

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	2 August 1996	Informal Technical Data

4. TITLE AND SUBTITLE
Canvas Knowledge Acquisition Guidebook, Version 1.0

5. FUNDING NUMBERS
F19628-93-C-0130

6. AUTHOR(S)
Mark Simos, Dean Allemang, Larry Levine, Organon Motives; Charles Hammons, Lisa Mantock, ScenPro, Inc.; Carol Klingler, Dick Creps, Lockheed Martin Tactical Defense Systems.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Lockheed Martin Tactical Defense Systems
9255 Wellington Road
Manassas, VA 20110-4121

8. PERFORMING ORGANIZATION REPORT NUMBER
Document Number
STARS-PA29-AC01/001/00

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of the Air Force
ESC/ENS
Hanscom AFB, MA 01731-2816

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AC01

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution “A”

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This guidebook describes the Canvas approach to systematic knowledge acquisition. Canvas synthesizes elements of two distinct methods: Scenario-Based Engineering Process (SEP) and Organization Domain Modeling (ODM). SEP provides knowledge acquisition methods and representation techniques. ODM provides a conceptual framework for data acquisition planning for the purposes of domain engineering. The guidebook incorporates extensive lessons learned from project experience in managing a large-scale knowledge acquisition effort in coordination with advanced technology development in the health-care domain.

14. SUBJECT TERMS

15. NUMBER OF PAGES
136

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

Data Reference: STARS-PA29-AC01/001/00
INFORMAL TECHNICAL REPORT
Canvas Knowledge Acquisition Guidebook
Version 1.0

Table of Contents

Data Reference: STARS-PA29-AC01/001/00
INFORMAL TECHNICAL REPORT
Canvas Knowledge Acquisition Guidebook
Version 1.0

Table of Contents

Table of Contents

Data Reference: STARS-PA29-AC01/001/00
INFORMAL TECHNICAL REPORT
Canvas Knowledge Acquisition Guidebook
Version 1.0

Table of Contents

Data Reference: STARS-PA29-AC01/001/00
INFORMAL TECHNICAL REPORT
Canvas Knowledge Acquisition Guidebook
Version 1.0

List of Exhibits

Prologue

This document is version 1.0 of the *Canvas Knowledge Acquisition Guidebook.* It proposes a new and innovative approach to planning and managing knowledge acquisition activities in engineering projects. The Canvas approach synthesizes concepts and techniques from:

- The Organization Domain Modeling (ODM) domain engineering method sponsored by the DARPA Software Technology for Adaptable, Reliable Systems (STARS) program, and
- The Scenario-based Engineering Process (SEP) method sponsored by the DARPA Healthcare Information Infrastructure (HIIP) program.

Canvas was developed collaboratively by the STARS and HIIP programs under the DARPA Noah's Ark initiative (see Section 2 for further background).

We intend to apply and evolve the Canvas approach in the future. We encourage trial use of Canvas and solicit reader review and comments as input to its future evolution. To learn more about Canvas, discuss how to obtain support in applying it, or submit feedback on the guidebook or your practical Canvas experiences, please contact one or more of the following people:

Mark Simos
Organon Motives, Inc.
One Williston Road, Suite 4
Belmont, MA 02178
Phone: (617) 484-3383 x11
Fax: (617) 383-3363
E-mail: mas@organon.com

Charles "Bud" Hammons
ScenPro, Inc.
2604 Spring Lake Drive
Richardson, TX 75082
Phone: (214) 307-7641
Fax: (214) 306-6818
E-mail: hammons@onramp.net

Dean Allemang
Organon Motives, Inc.
One Williston Road, Suite 4
Belmont, MA 02178
Phone: (617) 484-3383 x13
Fax: (617) 383-3363
E-mail: dta@organon.com

Dick Creps
Lockheed Martin
9255 Wellington Road
Manassas, VA 20110
Phone: (703) 367-1353
Fax: (703) 367-1389
E-mail: richard.e.creps@lmco.com

Thank you for your interest.

1.0 Document Overview

This guidebook describes the Canvas approach to systematic knowledge acquisition. Canvas synthesizes elements of two distinct methods: the Scenario-based Engineering Process (SEP) and Organization Domain Modeling (ODM). SEP provides knowledge acquisition methods and representation techniques, while ODM provides a conceptual framework for data acquisition planning for the purposes of domain engineering. The guidebook incorporates extensive lessons learned from particular project experience in managing large-scale knowledge acquisition efforts, and generalizes them to a method for managing knowledge acquisition that is compatible with the goals of both ODM and SEP.

At the heart of the Canvas approach is a framework for identifying and managing the difficult issues of the knowledge acquisition process. These issues include managing the interests of the various stakeholders to the knowledge acquisition enterprise (e.g., the people providing domain knowledge, the knowledge acquisition staff, and the targeted audience for the information gathered), managing access to written sources of information, and managing the record of the knowledge that has been acquired. Understanding these interests provides systematic ways of planning the knowledge acquisition effort to handle bias and potential sources of resistance, avoid inefficient use of experts' time, and flexibly coordinate knowledge acquisition processes to respond to shifting availability of resources and newly discovered information sources.

1.1 Purpose and Scope

Canvas is primarily concerned with planning knowledge acquisition activities in a project. This might make it seem as though Canvas therefore ends just as the project is beginning, but in fact, many of the most important decisions in a knowledge acquisition project are planning decisions, and these decisions continue to be made and updated throughout the lifetime of the project. The major planning decisions in Canvas are:

- selecting the knowledge acquisition staff (investigators) and knowledge sources (informants),
- determining the roles of these participants, as audience, information sources, customers, knowledge engineers, etc.,
- tracking and managing bias,
- managing variability in information resulting from differences of opinion, differences of viewpoint, or differences in work practice,
- managing access to source information, including problems of scheduling and confidentiality,
- managing access to new information, produced as part of the knowledge acquisition process, and
- selecting notations for representing the acquired knowledge.

Planning does not include the actual execution of any of the tasks planned above. In particular, this guidebook does not attempt to do the following:

- Provide detailed guidance in how to perform knowledge acquisition activities such as interviewing practitioners or examining artifacts.
- Compare Canvas with other approaches to data gathering or knowledge acquisition (although it may help the reader in drawing such comparisons).

All of these are potential topics for future papers and reports to complement this document.

The Canvas approach is directed mostly towards the problems of planning and managing a large-scale knowledge acquisition effort to support requirements engineering, domain engineering, or expert system development. However, many of the principles behind Canvas are applicable to any situation in which knowledge is being systematically elicited from one work practice to another. Canvas pays particular attention to the problems of transferring technical information, and is less well suited to more social studies of a culture, where the information being transferred is usually less technical. Nevertheless, many of the Canvas strategies for handling bias and responsibility are still applicable, even in such a setting. Although Canvas was designed from lessons learned managing a large project, many of the principles are just as applicable for small projects. Canvas is particularly appropriate in cases where a large number of (that is, more than two) professional cultures are involved in differing capacities, and the interactions among them must be carefully tracked. Canvas is also appropriate when the professional communities involved have complex structures; that is, they include several distinct groups with different stakes in the knowledge acquisition effort. A detailed description of the kinds of activities that qualify as KA from the Canvas perspective can be found in Section 3.1.

1.2 Intended Audience

Canvas emphasizes knowledge acquisition as an aid in understanding which technologies should be inserted into a workplace and how they should be inserted. We have attempted to write this guidebook from the "ground up", so that anyone who has given some thought to technology insertion issues can follow the concepts described here. However, the audience who stands to gain the most from reading this book consists of people who have had responsibility for placing new technology in a workplace, and would like to examine in more depth the reasons for success and failure of such attempts. Canvas challenges some assumptions about how the information about a workplace can be transmitted to the people who design systems for that workplace, and gives detailed advice about how a knowledge acquisition project should be organized, based upon these challenges.

More specifically, Canvas brings specific benefits to readers having one or more of the following organizational roles:

- *System/software project managers* — For project managers, Canvas shows the benefits of systematic knowledge acquisition, and the risks that one incurs when certain aspects are not taken into consideration. It also provides detailed advice about how to plan knowledge acquisition activities in a project and track knowledge acquisition workproducts.

- *System/software developers* — Canvas illustrates how the cultural differences in workplaces have an impact on the information needed for effective requirements engineering. Thus system developers will gain an appreciation of the potential value to be gained from knowledge acquisition as an integral managed part of technology development. They will better understand why KA is a distinct discipline and the importance of having the work performed by competent practitioners.

- *Experienced knowledge engineers* — Canvas puts well-known knowledge acquisition techniques (interviewing, verification, representation) into a general framework for application on large projects. Knowledge engineers will see how their efforts are related to other KA activities. Depending on their level of experience, they might find new insights into the many aspects of KA treated by Canvas, including bias, variability, the specific challenges posed by performing KA in technical settings, or acquiring knowledge about multiple systems.

- *New knowledge engineers* — Canvas provides a concise description of knowledge acquisition, the assumptions behind it, its goals and the infrastructure needed to support it. These items can be of great help in learning what is expected of a knowledge engineer.
- *Domain modelers* — Canvas is derived in part from the ODM domain engineering method and is thus directly applicable to the data acquisition needs of domain modeling, especially of the sort described by ODM.
- *Social scientists* — Experienced knowledge acquisition people from a social sciences, academic or humanist background will gain better understanding for the distinctive challenges and opportunities of performing KA in a technology-rich environment. This will become of increasing importance for KA efforts as technology becomes a more ubiquitous presence in most cultural settings.
- *Knowledge sources/informants* — Anyone who will potentially be in the role of informants or knowledge sources for a KA effort will better understand the process in which they are taking part. This will enable them to take a more pro-active and collaborative role, perhaps including observation of instances of investigator bias and articulation of these concerns.
- *Support technology developers* — Developers planning to provide supporting tools for knowledge acquisition will find detailed recommendations for how a repository of information must be managed in a KA project, as well as requirements for managing the other resources in the project.
- *Students, researchers, etc.* — Anyone who is interested in knowledge acquisition or in using packaged knowledge can use parts of this document to understand how knowledge is organized and why, making knowledge acquisition and use more effective.

For all readers, Canvas offers an awareness of how culture plays a role in knowledge acquisition and what can be done to address cultural issues.

1.3 Guidebook Organization

The body of the guidebook is organized into the following sections:

— *Section 1: Document Overview (this section)* — Defines the purpose, scope, intended audience, and organization of the document and offers guidance in how to read it.

— *Section 2: Canvas Overview* — Describes the background and motivation for the project that developed Canvas, the context in which it was developed, and an overview of the knowledge acquisition planning process which is the scope addressed by the method.

— *Section 3: Canvas Core Concepts* — Articulates the conceptual foundations of Canvas, defining the boundaries around what is considered knowledge acquisition (for the purposes of this guidebook), and defining the basic terms used and implications of these foundations.

— *Section 4: Planning the KA Enterprise* — Provides a set of basic, practical guidelines for using Canvas to plan and manage a knowledge acquisition project.

— *Section 5: Representation of Knowledge* — Offers criteria for comparing different notations used for representation and how these notations impact the knowledge acquisition process. Section 5 also provides a list of sample notations that can be used in knowledge representation.

— *Section 6: Dossier Planning and Management* — Provides detailed advice about how to structure a repository (herein called a "dossier") of knowledge acquisition products, including "starter sets" to help initialize the dossier.

— *Section 7: Conclusions* — Reviews the key principles embodied in Canvas and proposes possibilities for further work.

The guidebook also includes the following supplementary material:

- *Appendix A: Canvas as an ODM Supporting Method* — Shows how Canvas fits into the ODM process framework and can be used as an ODM "supporting method."
- *Appendix B: Representing the Knowledge Acquisition Process* — Uses Canvas on itself; the knowledge acquisition process is shown, represented in notations commonly used in Canvas.
- *Appendix C: Canvas Lexicon* — Definitions of the terms that are fundamental to understanding Canvas.
- *Bibliography* — Bibliographic entries for documents referenced in the guidebook and other documents related to Canvas.

1.4 How to Read the Guidebook

Each segment of the audience has different needs and interests and will thus benefit most from different portions of the guidebook. All segments of the audience should read Sections 2 and 3 (consulting the Canvas lexicon in Appendix C as needed) to gain a basic understanding of the method.

Readers with a background or interest in domain modeling will find Appendix A of particular interest. Developers of technology to support knowledge acquisition efforts will find Section 6.0 and Appendix B to be of particular interest. Knowledge engineers are most likely to find Section 5.0 to be most useful for their needs, while project managers will need to master the ideas in Section 4.0.

The name and number of the current section is included in the header on each page to help the reader navigate through the document.

Treatment of key Canvas terms: In general, when a key term is first introduced in the guidebook, it appears in a ***bold italic*** typeface and is defined at that point. Such terms are also sometimes shown in bold italic when first appearing within other sections of the document to reestablish their identity as lexicon terms. All such terms also appear, with definitions, in the Canvas lexicon in Appendix C so they can be easily looked up whenever they are encountered in the document. Appendix B shows how many of the core terms relate to one another, and is itself an interesting case study in knowledge representation.

2.0 Canvas Overview

People acquire knowledge informally on a regular basis, as a natural outgrowth of performing work tasks, reflecting on past experience, communicating experiences to others and learning from others' experiences to improve practice. In many situations, however, it becomes important to follow a more systematic knowledge acquisition process. Familiar examples are rules for acquiring evidence in trial law or procedures used by journalists in gathering background information. The notion of knowledge acquisition as a distinct phase of a technology development effort has gained greatest prominence in the expert systems/artificial intelligence (AI) field, where the goal has generally been to codify expert knowledge in some domain into a representation that can serve as a basis for automated deduction and decision support. Knowledge acquisition is becoming more and more important in large software projects in which systems will be built to interact with many extant systems and work practices, in which comprehensive information about these interactions is imperative for the proper integration of the software to be produced. In particular, the approach to knowledge acquisition presented here has been motivated to a great extent by the needs of large-scale software domain engineering efforts.

Systematic knowledge acquisition involves repeatable procedures for making key decisions in planning and performing knowledge acquisition, and for recording results of KA activities in a way that preserves essential contextual information about the data acquired. A systematic approach to knowledge acquisition addresses the following issues (among others):

- *Sources of the information*: Where did the information come from and how was it obtained? Was there possible bias, misinterpretation or selective filtering on the part of the person who obtained the information?
- *Handling multiple information sources*: What is the relative convergence or divergence of opinion in some professional community? Does this represent variance in opinion and belief or a natural variance in the phenomena described? For example, if three medics describe a standard procedure and give three contrasting sequences of tasks, are two of them wrong? Are they describing particular events, "routinized" generic descriptions, or the official textbook version of what's supposed to take place in the given situation? Or are there different practices involved because the three medics work in different units, different hospitals, different states?
- *Managing the data acquisition process*: This includes issues such as budgeting resources for data acquisition. Given scarce resources, unpredictability in accessing experts or the effort required for specific sessions, what are the most important information sources to consult? How do we know when we have acquired enough data? Handling of potentially sensitive or proprietary data must also be considered.
- *Access to the information.* Who will be using the information gathered, and how? Are there many audiences with differing perceptions about what the knowledge should be and how it should be used? How can we help all of these audiences find the information they need?

This guidebook presents an initial framework for planning and managing a systematic knowledge acquisition effort. The framework is presented as a core set of concepts, planning guidelines, example representations, and a set of guidelines for initializing a ***dossier*** — a repository containing knowledge acquisition workproducts organized to facilitate access by the target audience.

We call the framework presented in this document Canvas. One of the meanings of the English word "canvas" is:

> A coarse cloth of open mesh weave on which embroidery or tapestry is done.

The name of the Canvas framework, as will become clear later, is intended to evoke an image of weaving strands together into a coherent tapestry of knowledge. The strands begin as independent threads of information, which are guided by Canvas to form a woven fabric that has more structure and meaning than the threads alone. We will use this metaphor throughout the guidebook when we speak of the interactions between the life cycles of the various elements that participate in the knowledge acquisition process.

We will refer to Canvas in different contexts as an "approach" (where the emphasis is on the key concepts and concerns/issues raised) or as a "framework" (where the emphasis is on the core set of terms and elements and their semantic relations). We do not consider the current document to be a stand-alone method or process model. By focusing on knowledge acquisition *planning* we have excluded much specific detail about how to carry out knowledge acquisition activities such as interviewing or analysis. These are beyond the scope of the current document, but might well be expected as part of a comprehensive method for knowledge acquisition. In addition, while this guidebook reflects real project experience (e.g., the SEP-based TCIMS program, ODM pilot project efforts), it does not represent a distinct process that has been followed multiple times in order to yield a repeatable process description.

2.1 Project Background

This guidebook is the result of a Defense Advanced Research Projects Agency (DARPA)-funded task involving collaboration between the Software Technology for Adaptable Reliable Systems (STARS) and the Health-care Information Infrastructure Program (HIIP) Programs. STARS participating organizations included Lockheed Martin Tactical Defense Systems, Organon Motives, Inc., and WPL Laboratories, Inc. HIIP participating organizations included ScenPro, Inc. and the University of Texas/Arlington. The overall objective of the joint task was to develop a software engineering method which integrates key elements of the STARS ODM and HIIP SEP methods. These methods are described briefly below, followed by additional rationale and motivation for the integration activity.

2.1.1 ODM Background

Organization Domain Modeling (ODM) is a highly tailorable and configurable domain engineering method, useful for diverse organizations and domains, and amenable to integration with a variety of software engineering processes, methods and implementation technologies. The method offers a systematic, exemplar-based approach to analysis of commonality and variability, specifically addressing analysis of both legacy systems and requirements for new systems to derive reusable assets focused within a particular domain. ODM grounds the domain modeling in the context of the organization and the relevant stakeholders. Key features of ODM include the following:

- In addition to focusing on the strictly technical aspects of domain engineering, ODM emphasizes analysis of the diverse stakeholders that form the organization context within which each domain engineering effort is conducted.
- ODM provides systematic techniques for identifying and selecting highly focused domains of

strategic interest within larger business areas, and for incremental and iterative scoping to mitigate risk and produce robust, coherent domain models.

- The ODM modeling life cycle details the transformation from descriptive modeling of legacy systems, artifacts and experience to prescriptive specification of architecturally integrated assets, designed for a well-scoped range of variability and characterized in terms of features relevant to domain practitioners. The resulting domain model offers a sound basis for making the design decisions and trade-offs required to engineer sets of reusable components that are robust, high quality and natural to use.

ODM domain engineers study a carefully selected set of representative software systems within a domain. ODM provides criteria for selecting from a wide variety of knowledge elicitation techniques, including artifact analysis, interviewing of domain informants, and process observation. This might include analysis of artifacts from across the software life cycle, as well as use of ethnographic techniques such as those described in [40] and [41] (e.g., to capture process knowledge and undocumented "techlore" of application developers, users and other domain stakeholders). The resulting data is formalized in a domain model which represents common and variant features of systems in the domain. The process requires a unique mix of technical, conceptual and organizational skills on the part of the domain modeler.

Under funding by the DARPA STARS Program, ODM has been extensively documented in a guidebook [46] as well as in shorter papers [36] [37]. The guidebook provides explanations of key concepts, a formal process model (documented in $IDEF_0$ process modeling notation [18] [22] [39]), work product descriptions and templates, and detailed, practical guidelines for domain engineering projects.

2.1.2 SEP Background

The Scenario-based Engineering Process (SEP) [12] is a user-focused methodology for system development. The methodology has been applied and is currently in use in various health care programs, including the DARPA Trauma Care Information Management System (TCIMS) program. SEP has strengths in engaging the user in all phases of a project, and benefits from a well-structured approach to eliciting information, utilizing scenarios as a means of engaging the user.

As shown in Exhibit 1, SEP consists of three processes: knowledge acquisition (KA), knowledge engineering (KE), and system engineering (SE), which form a well-defined semi-ordered set of procedures.The result of a SEP process is the construction of a component-based architected system.

SEP has a number of objectives that distinguish it from other system development methodologies. These objectives are to improve communications among project stakeholders, facilitate the assignment of responsibilities, and maintain traceability within interdisciplinary system-in-the-large development efforts. In order to achieve these objectives, SEP focuses on scenarios for knowledge acquisition and validation, iterative "build-a-little, test-a-little" prototyping, and a strategically planned incremental development approach. For the purposes of knowledge acquisition, and hence for Canvas, SEP's focus on scenarios is of the greatest relevance

In SEP, a scenario is a single path through some work process, similar to case studies familiar from medical practice. In fact, in the TCIMS project, where SEP was applied to a medical domain, many of the scenarios read very much like case studies. Scenarios are the key components driving communication and traceability in SEP. Scenarios participate in the SEP lifecycle at several points. During knowledge acquisition, they are used to communicate with the prospective users of the technology, to determine their requirements and derive system specifications. Filtered

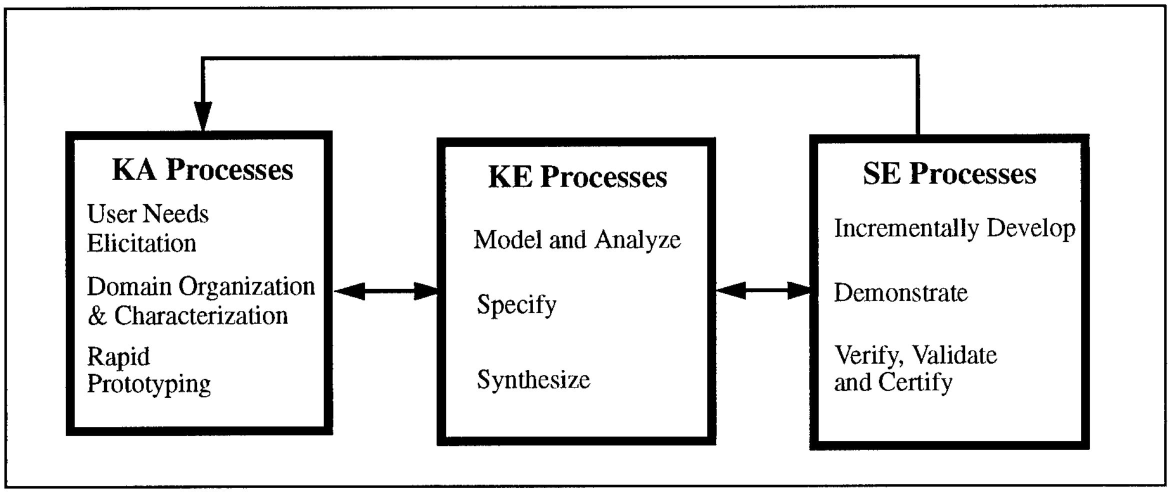

Exhibit 1. Scenario-based Engineering Process Overview

and generalized scenarios result in task analyses, which, along with other knowledge acquisition products, are used during knowledge engineering for object-oriented design and architecture synthesis. Finally, the scenarios themselves are used again during testing and validation stages of system engineering to evaluate the developed system.

The ramifications of using scenarios as the basis of the engineering process are subtle but important. Scenarios are directly understandable to the prospective users, since they are simply records of their experiences; on the other hand, the same scenarios can be used by system developers as an "acid test" for software products, including specifications, requirements, and executing code. Using the same scenarios throughout the process facilitates traceability of workproducts. Since the scenarios are relevant to both the users and the developers (although in different capacities), they foster communication between the two groups. As opposed to other descriptions of the workplace (e.g., object descriptions, task structures, organizational charts etc.), scenarios provide direct information about how the practitioners in some workplace interact with other practitioners or systems, thus facilitating the identification of which practitioners or subsystems are responsible for which actions.

This final point is a critical aspect of the underlying motivation for SEP. SEP recognizes that the interactions of prospective users with their environment is flexible and dynamic. Static generalizations that talk about these interactions are useful for specifying and building systems, but in the final analysis, the systems must be responsive to the detailed, dynamic nature of the interactions. By basing the process on scenarios, this accountability is retained, since the scenarios record the interactions themselves in their raw form, without generalization. This ability to attend to the interaction between the end users and their environment is the primary value that SEP adds over other system development methodologies.

2.1.3 Motivation for the Canvas Method

Both the ODM and SEP methods are rooted in conventional software engineering approaches. Each method has diverged from those conventional approaches in significant, although different, ways. Yet they share many of the same concerns and offer complementary perspectives on key software engineering problems. There are thus substantial opportunities for synergy among the two methods.

As the SEP method summary above implies, SEP focuses primarily on engineering individual systems. Although workproducts produced in developing each system may prove of value in developing subsequent systems, the method does not emphasize multi-system analysis with the explicit objective of developing components that can reused in multiple application contexts. In contrast, ODM, as shown in Exhibit 2, is one of a class of methods called *domain engineering*, which extends conventional software engineering approaches by focusing explicitly on analysis across multiple application contexts to support systematic reuse. While many domain engineering approaches emphasize the exploitation of commonalities across systems, a key feature of the ODM domain engineering approach is its rigorous analysis of variability across systems to support systematic management of alternatives within a domain. This rigorous treatment of variability is one of the key differences between ODM and SEP.

Within their respective milieus (system and domain engineering), SEP and ODM both rely heavily on knowledge acquisition and employ a variety of knowledge acquisition techniques. The primary difference between the methods in this area is the community of practice which they each emphasize. SEP focuses primarily on acquiring knowledge about the environment in which domain practitioners (i.e., end users) are practicing their craft. This is done mainly through scenario-based interviewing and related techniques. The knowledge acquired in this way helps system developers to get a clear understanding of how work is done in that environment and gain insights into how it can be better automated. ODM, on the other hand, places a stronger emphasis on acquiring knowledge about the environment (or setting, in ODM terms) in which applications are developed. It focuses less on interviews with practitioners (i.e., developers) and more on analysis of existing system artifacts in the domain.

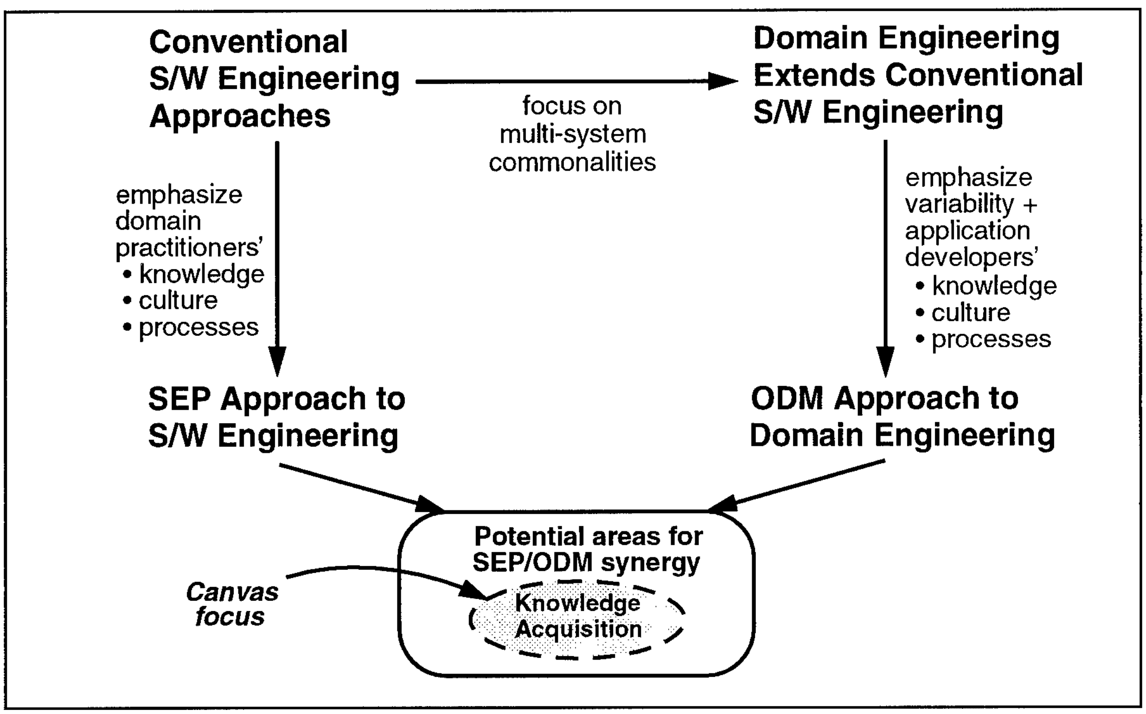

Exhibit 2. Motivation for a Composite ODM/SEP Method

In comparing SEP and ODM, it became clear that there were several areas in which the methods complemented one another and could benefit synergistically from cross-fertilization. These included modeling approaches and architectural concepts, among others. However, the area offering the richest integration opportunity was knowledge acquisition. Each method could benefit from a knowledge acquisition approach that unified the SEP and ODM perspectives (e.g., single

system versus multi-system; domain practitioners versus application developers; scenarios/interviews versus artifact analysis.) It was from these insights that the Canvas approach was born.

The Canvas knowledge acquisition approach is based not only on study and comparison of the SEP and ODM methods themselves, but on the experiences gained by applying them in a variety of contexts. Canvas, in recognition of these lessons learned, is designed to be usable in a number of knowledge acquisition settings, including but not limited to SEP and ODM. Canvas can be used as part of any SEP or ODM effort, but the Canvas principles (even though derived from SEP and ODM) are applicable in many other contexts. This document has therefore been written not to assume that Canvas knowledge acquisition activities are part of a larger SEP- or ODM-based effort. We hope that a wide range of projects that rely on knowledge acquisition can thereby benefit from our experiences in developing and using SEP and ODM and integrating their perspectives in the Canvas framework.

2.2 Canvas Goals

Canvas has been specifically designed in response to the knowledge acquisition needs that have been perceived in ODM and SEP. A number of knowledge acquisition methods satisfy many of these needs. We have set the goals of Canvas to respond to particular needs that, in our experience using these two methods, have not been fully met or adequately treated by other methods. These goals include:

- *Make the study of cultural knowledge central to KA.* Both ODM and SEP offer, as added value over competitive methods, a particular attention to the interaction of prospective users with their environments, including automated systems and other practitioners (possibly from different backgrounds). This attention places specific demands on the knowledge acquisition process, in that it must enable acquiring knowledge about interactions among people with widely varying work backgrounds and/or systems designed by such people. One goal of Canvas is, therefore, to make the study of interaction between different work settings central to the knowledge acquisition process.

- *Foster utilization of the full spectrum of available knowledge sources.* Traditionally, knowledge acquisition practices have focused primarily on collecting information via interviews with informants in domain practitioners' settings ("usage settings," in ODM terms). Artifacts or documentation are used more cautiously, as supplemental material or background briefings for interviewers. There is a sense that we will get the "real story" from interviews whereas we might get only the official story (policies, regulations, etc.) from documentary sources. Study of artifacts could be used more effectively in these settings via walk-throughs, validation of interview data, etc.

  Conversely, there has been a tendency to under-utilize observation of or interviews with people in developers' settings. Those who build the systems have a unique picture of the work domain in which the systems will operate. This picture is distinct from the practitioners' view, but it has a powerful effect on the kinds of systems built; also, once those systems are fielded, the developers' picture will have powerful impact on the work practice itself.

  Acquiring knowledge by cross-checking informant interviews, analyzing artifacts, and directly observing work practice provides a richer set of data and more possibilities for robust validation. Another goal of Canvas is, therefore, to support an integrated framework where different types of knowledge sources can be considered as an integrated whole, not as a grab-bag of distinct types of data.

- *Emphasize legacy systems and anticipated new systems.* In the DoD arena there are many environments where introduction of new systems or technologies must be preceded by some

understanding of the large number of existing "stovepipe" legacy systems. Many of these do not currently communicate or interoperate to any acceptable level. Knowledge acquisition to support technology development in these environments must have systematic ways of accounting both for anticipated new systems and existing legacy systems in the environments studied. In domain engineering, study of legacy systems is integral to the task, although the typical objectives are comparative analysis to support reengineering for reuse, rather than analysis to support introduction of new systems that can interoperate with or link legacy systems. From the KA standpoint many of the challenges are similar. A goal of Canvas is to extend traditional KA techniques to understand the role of legacy systems when addressing issues of technology development and usage.

- *Validate/increase credibility of KA within technology community.* An important lesson that has been learned in applying SEP is that the people who were responsible for developing systems based on the results of knowledge acquisition efforts were not always convinced of the value of performing knowledge acquisition in a systematic way. We have therefore made it a goal of the Canvas effort to determine the causes of this reluctance, and to design aspects of the process that will encourage the transfer and use of knowledge acquisition results.

2.3 Validation through TCIMS Experience

In developing the Canvas method, we have drawn extensively on the project experience of SEP method providers and knowledge acquisition specialists in the Trauma Care Information Management System (TCIMS) project. The following paragraphs provide some context for this project, from which examples are cited throughout the guidebook.

The Trauma Care Information Management System (TCIMS) project is a two-year effort to demonstrate radical improvements to the nation's Trauma Care Information Infrastructure, both civilian and military. A consortium of 13 organizations under DARPA guidance is designing, and will demonstrate the benefits of, advanced computing and communications solutions to accomplish this goal. In pursuit of this goal, the TCIMS project has the following objectives:

- The completed system will be commercially self-supporting. Consortium members intend to develop and produce products that conform to the TCIMS reference architecture and are commercially viable in both military and civilian use, both urban and rural.
- TCIMS will improve field trauma care by providing relevant medical and patient information to field medical care providers, and transmitting patient data ahead to the hospital to minimize delays in scheduling emergency facilities and resources.
- The TCIMS consortium will develop national-level trauma care information management standards leading to rapid price reductions in the cost of such information and inter-operability among all users and providers.
- The consortium is developing a TCIMS architecture to promote continuing trauma care system development and innovation.

When medical care arrives at the scene of a medical trauma scene, a Personal Status Monitor (PSM) might be attached to the ill or injured person to sense their condition. In a military environment, each patient is expected to have an identification and information card such as the AT&T "Smart Card," which provides patient identification and a minimal medical history. The PSM will have a micro controller/memory module and a communication scheme. It will monitor the patient's vital signs and location, and will give an alert if the patient's condition becomes critical. The PSM is being developed by another DARPA program.

PSM and Smart Card data would be read by a Field Medic Associate (FMA) computer carried by the medic, which will provide a multimedia display of patient information. Voice command is among the interfaces being considered to input information to the computer. The Field Medic Associate will also provide the field medic with access to clinical knowledge, information, and advice needed to treat each patient. The FMA will allow the medic to gain on-line information about resource and facility availability.

The individual Field Medic Associates will feed information into a Field Medic Coordinator computer, which will give a situation overview for crisis managers and aid in focusing attention on the most critical patients. The Field Medic Coordinator computer will in turn feed information into the trauma center's Trauma Center Coordinator computer in the Clinical Workstation System, so that a complete and current case history on each patient will be built at the hospital even as the patient is being treated in the field. That history will follow the patient through the trauma care system. The Trauma Center Coordinator will provide rapid, complete information to the Patient Care Units and to portable Trauma Care Associate computers carried by care providers in the hospital. The entire TCIMS will allow direct communication among medical decision-makers.

TCIMS will, therefore, provide rapid treatment procedures to help the medic make treatment decisions in complex and confused situations. The medical center will have access to on-line patient records, including x-rays and other test results. TCIMS will reduce time to treatment, will bring knowledge and information to decision-makers, will increase ratio of treatment to processing, and will reduce overall patient processing time. Finally, TCIMS will link clinical, tactical, and strategic planning and management.

Role of TCIMS data in preparing this guidebook. The opportunity to incorporate lessons learned from the extensive KA work undertaken as part of the TCIMS project has enriched the Canvas framework and ensured that it is grounded in experience in (at least) health-care related domains. In the process of developing Canvas, the authors studied the TCIMS Knowledge Acquisition Plan for TCIMS in detail and debriefed Lisa Mantock, one of the lead knowledge acquisition specialists on the project, on the planning process. We also observed a TCIMS-related interview session with a helicopter pilot involved in medical emergency transport, and worked with scenario data, domain models represented in the Loom knowledge representation system, and an on-line repository of TCIMS knowledge acquisition data (SEPWeb). Most of these materials are proprietary to the TCIMS consortium members and thus have not been cited in the Bibliography section or included directly in examples within the text. Nevertheless, the material provided one extensive data point for the framework and related recommendations offered in this guidebook.

3.0 Canvas Core Concepts

This section introduces core concepts necessary to understand the Canvas framework for planning knowledge acquisition and managing the results of the KA process. Section 3.1 provides an explicit set of defining features for the general phenomenon we call knowledge acquisition. Section 3.2 introduces the key elements of the knowledge acquisition "canvas" as a central metaphor for the approach in this guidebook. Section 3.3 extends this basic framework to address issues such as knowledge acquisition in technology-intensive settings and the management of variability. These issues are essential for systematic knowledge acquisition that is to be integrated with system or domain engineering projects.

3.1 What is Knowledge Acquisition?

We view knowledge acquisition as a special case of a much broader area of human activity we will refer to simply as ***knowledge creation***, any process that results in new knowledge being created. Individual learning, formal teaching, process capture, research, and knowledge acquisition (the focus of this document) are all forms of knowledge creation.

To differentiate knowledge acquisition from other forms of knowledge creation we will need to define a few basic concepts. We will use the general term ***work setting*** to denote any environment where people interact with each other and regularly perform work processes. The notion of a work setting implies a certain stability, in that the same people work together on a routine or regular basis. These ***practitioners***, in the context of the various work settings (or more descriptively, work practice settings) where they interact, form a ***community of practice*** (often referred to in this document simply as a "community".)

For the purposes of this document, we characterize ***knowledge acquisition*** (hereafter abbreviated as KA) as a knowledge creation process that involves the transfer of knowledge across communities of practice. Although we deliberately leave the base term "knowledge" itself undefined in our lexicon, the approach here is roughly aligned with the concepts of knowledge and learning from the "learning organization" field; i.e., knowledge as the capacity for effective action in a given domain, where effectiveness is assessed by a community of practitioners (see [32], [17]).

We are particularly interested in knowledge about work practice that may be ***embedded*** in a work setting, that is, hidden behind tacit or unarticulated assumptions. This kind of knowledge is often transferred informally to new practitioners within the setting; numerous problems arise in trying to transfer such knowledge to new work settings or communities of practice. Many formal processes and techniques in knowledge acquisition are designed to deal with this specific set of challenges. The Canvas approach is an extension of some of these more systematic approaches to KA to accommodate special requirements of KA for system and domain engineering. The general relationships between these areas is depicted in Exhibit 3.

The following sub-sections further elaborate our provisional definition of knowledge acquisition. Section 3.1.1 outlines the defining features of KA; Section 3.1.2 clarifies the distinctive aspects of KA by comparing it to other, more familiar forms of knowledge creation; Section 3.1.3 introduces some high-level "modes" of KA in terms of the primary objectives of the activity and audience for the workproducts produced.

3.1.1 Defining Features of KA

Knowledge acquisition is a process involving at least two communities of practice: a ***focus community*** and an ***investigator community***. The initiating condition for a knowledge acquisition

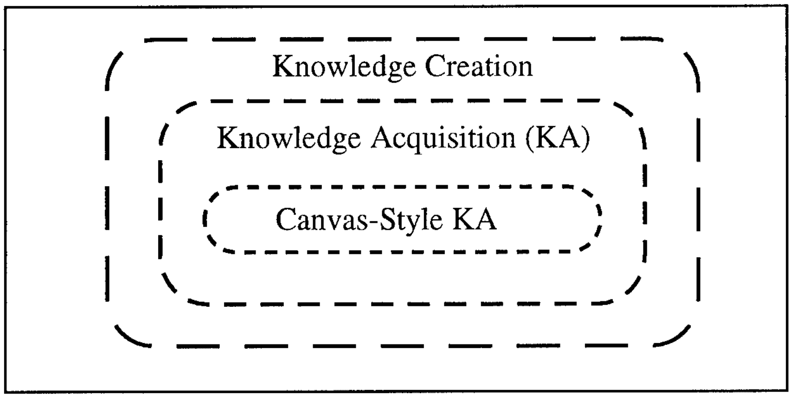

Exhibit 3. KA as a Special Type of Knowledge Creation

effort is that some knowledge (the ***focus of interest***) embedded in the work practice of the focus community is of interest to some stakeholders in the investigator community. The beneficiary of the knowledge acquisition process is another community, called the ***target community***, which might be the focus or the investigator community, or perhaps a third distinct community. (Relations between these communities are discussed further in Section 3.1.3. The important concept here are the distinct *roles* played by each community.)

Three distinct elements qualify a process as knowledge acquisition; the knowledge must be ***elicited***, ***codified*** and ***transferred***.

- Knowledge is ***elicited*** from a ***knowledge source*** drawn from the focus community. This elicitation process involves the active participation of a practitioner from the investigator community, an ***investigator***. After elicitation, the investigator has some knowledge at his or her disposal that he did not have before the elicitation.

- Some of this elicited knowledge is ***codified*** in a knowledge acquisition workproduct. For example, if an expert has been interviewed then some interview report must be written. A video or audio tape may have been made as well. If a design document has been studied as part of the KA effort, then some notes on the relevance of the document to the topic of interest will be written by the investigator. These often informally created workproducts are in fact an essential part of the knowledge acquisition process, for without them the information transfer reduces to the learning that has taken place for the investigator. This learning in itself is not knowledge acquisition. All workproducts created during the knowledge acquisition process will be stored in the ***dossier***.

  Every workproduct created in this way has an intended ***audience***, a community of practice for whom it is intended; most KA workproducts will have the target community as audience. Knowledge has been effectively codified into a workproduct if a member of its intended audience can learn the knowledge by examining (reading, viewing, running, etc.) the workproduct.

- In knowledge acquisition, knowledge is ***transferred***, not just among individuals, and not just from expert to novice individuals within a given community, but across communities. This is a key defining characteristic of the process, and distinguishes knowledge acquisition from a number of related activities. This transfer activity is not one of the routine work practice activities of the focus community (although it may conceivably be such for the investigator community). The target community to which knowledge is to be transferred can be either the focus community, the investigator community, or a distinct third community.

Each one of these three elements— elicitation, codification and transfer— must be present for an activity to qualify as a knowledge acquisition activity in the sense used in this document. This definition is intended to scope what we mean when we refer to KA.

3.1.2 Distinctive Aspects of KA

There are two major forms of knowledge elicitation:

- The investigator may meet with a practitioner from the focus community, such as an expert in a particular field. In such an ***interaction*** the person serving as the knowledge source is termed the ***informant***. In order for this interaction to qualify as knowledge elicitation in the above definition, investigators will gain some knowledge about the topics of interest as a result of the interaction.

  Similarly, informants may learn new things about their own topics of knowledge as a result of the interaction. This may occur through the act of reflection required to articulate terms and concepts to the investigator, who lacks some of the implicit context of other practitioners from the informant's community. Or the informant may be brought into contact with other people from the same community that he or she would not have encountered as part of routine work activities; and this encounter may engender new knowledge as well. In any case, the new learning created for the informant is also part of the knowledge elicitation event, but not a defining aspect.

- Investigators may also perform a ***study*** of some workproduct drawn from a work setting of the focus community. For example, they might read a user manual, a field report, an article in a newspaper, etc. A workproduct that is studied for the purposes of knowledge acquisition is called an ***artifact***. Since we are particularly interested in knowledge that is tacitly embedded in the focus community, the artifact should play a distinct role in the focus community. Typically the use of the workproduct as an artifact is different from the original purpose for which the workproduct was created; it is studied in a different context; the knowledge acquisition context.

Observation is a third form of elicitation, but is not at the level of importance for our purposes as interaction or study. Observation is most useful as a check of the other two forms of elicitation or in settings where there are few artifacts to study and/or informants do not easily articulate the nature of their work. Also, one brief observation very early in the KA planning process can be extremely helpful in planning what artifacts to study and who to interview and how to interview them.

In order to clarify some of the subtleties of the definition we use above, we will examine a number of learning activities that bear close similarity to KA, but fail to match the above definition in one or more illustrative ways.

- *Information Transfer.* Within a work practice setting, people routinely exchange information in order to get their jobs done. For example, a salesman takes an order from a customer and transmits it to the stocking department. Information has been transferred between people, but it is to facilitate the routine work activities that are the focus of the work practice setting. Such information transfer fails to qualify as KA because it is confined within a single work setting, and because the transfer is part of the routine practice in that setting.

- *Experience.* As people perform repeated tasks in a given work setting, they gain experience with that work. Exposure to repeated situations and variations in the conditions encountered increases their competence over time. Though they are "acquiring knowledge" through experience this kind of learning is also, strictly speaking, out of the scope of knowledge acquisi-

tion as we define it here. This experience is not yet transferable to others; nor has it been codified in any formal way.

- *Reflection.* As people gain experience, they often actively reflect on that experience and create new knowledge through this process of reflection.The reflective person gains knowledge through an activity that is often not part of the normal work practice, involving remembering and reviewing activities outside of the work context, or just a brief period of "staring out the window" time. Reflection does not qualify as KA, since the knowledge has not been transferred, nor even, more importantly, *rendered more transferable.* The new knowledge still affects the individual, not their overall community of practice. No attempt has been made to cross from one community of practice to another.

- *Writing Down.* As reflection and experience become more formal people will write down or codify their knowledge. The act of "writing it down" is an essential element of knowledge creation. The codification could involve something as simple as a checklist of items to remember each time a process is performed, something more formal such as a procedures manual, or the automation of the process through a software application. When the write-up is intended for personal use, or for the use of professional colleagues within the same community of practice, the activity is not KA, since no effort has been made to transfer from one community to another. Explicit attempts to write up material for other work settings may very well qualify as knowledge acquisition, though only if elicitation is consciously performed. Such attempts made by members of the focus community are fraught with problems of bias.

- *Learning.* The processes described above—work practice, repeated experience, reflection, codification—are all aspects of individual learning. Once knowledge has been codified, e.g., in textbooks, etc. it can be effectively transferred among individuals. When new people are brought into a work practice setting, they generally go through an extensive learning process to become effective practitioners. The knowledge has been effectively elicited, codified and transferred. But since the readers are being brought into the same community of practice as the textbook authors, no cross-community transfer has been made.

- *Transfer within a Community of Practice.* If an expert in a field has a meeting with colleagues, describing some new idea, the transfer of knowledge remains from individual to individual. If the expert writes a report on the idea for her colleagues, the knowledge is thereby made accessible to other members of the expert's community of practice. This activity fails to be KA because the knowledge has not been made available to a *new* community of practice.

- *Transfer across Communities of Practice.* If an interested party from some other community of practice reads a book that is accepted as an authoritative source, and learns these ideas, then knowledge has been transferred from one community to another. This activity is commonly mistaken for KA; however, the ideas have not been made more accessible to the new community in general, only to a single member of that community. A similar situation holds if the interested party talks to the expert or attends a lecture, and thereby learns the ideas. Personal transfer can be made by pursuing a personal change to a single member of the community; transfer to an entire community involves finding a way to present the knowledge in a way that is accessible to members of the new community in general. If the interested party accomplishes this, then knowledge acquisition has indeed occurred.

Each of these learning activities is a valid and useful activity in its own right, but does not qualify as knowledge acquisition in the Canvas context unless the three elements — elicitation, codification and transfer — are present.

3.1.3 Modes of Knowledge Acquisition

Given the definition of KA above, we can identify several distinct modes of knowledge acquisition. Each mode corresponds to one of the three basic configurations of the knowledge acquisition process shown in Exhibit 4.In all three configurations, the elicitation process involves the focus community as a knowledge source and results in some knowledge creation in the investigator community. The configurations differ based on the intended audience (i.e., target community) of the workproducts produced by codifying the elicited knowledge:

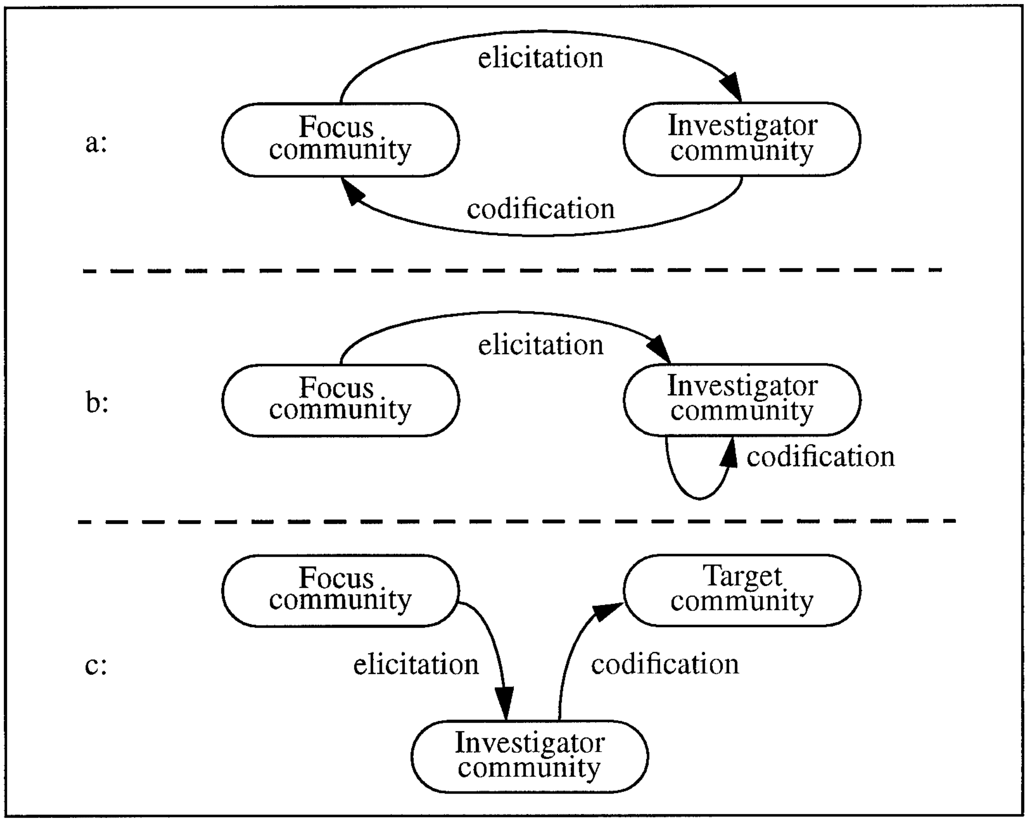

Exhibit 4. Basic KA Interactions among Communities of Practice

- *Audience is the focus community (configuration A):* The knowledge acquisition effort organizes material for the benefit of the focus community. One objective of knowledge acquisition can be to clarify knowledge for a professional community. An example would be a medical database, designed to allow doctors from one specialty to access new results from some other but related specialty. Many communities perform this function through their own research; in this case, an external investigator community brings its experience in knowledge organization to such a collaboration. Another example is the creation of standards, by standards committees. The process draws upon experts from the domain to be standardized, as well as someone who knows the subtleties of the trade-offs involved in writing usable standards. In such cases, the audience for the workproducts is the focus community.
- *Audience is the investigator community (configuration B):* The knowledge acquisition effort results in increased understanding for members of the investigator community. One common objective of knowledge acquisition is when one community studies another. Academic studies of a work practice are the simplest example of this mode of knowledge acquisition. It is

also the standard pattern for requirements engineering, where a software analyst or developer studies a work setting to determine requirements for a new system. This can happen as well in any situation in which a new interaction between technologies requires that one community learn the basics of another. Ergonomic studies are typical examples, where information about the work setting into which a technology is to be place must be acquired.

- *Audience is a third, separate target community (configuration C):* In this case, the target community for the workproducts is distinct from the focus and investigator communities; thus, the knowledge acquisition effort plays an intermediary, "bridging" role between the focus and target communities. In circumstances in which practitioners in the focus community exercise a great deal of autonomy and discretion, it is often necessary to utilize an investigator community that specializes in knowledge acquisition. Often, the investigator community includes practitioners with backgrounds from both the focus and target communities, which makes them particularly adept at "bridging the gap." The TCIMS project was organized in this way, with a separate team whose responsibilities were focused on knowledge acquisition.

Conflicting stakeholder interests in a large knowledge acquisition effort will result in multiple audiences, each with a particular focus of interest. For example, whenever elicited knowledge is recorded or codified in a way intended to facilitate review and validation of that data by informants in the focus community, that information will be of potential direct interest to those informants, as practitioners, beyond its validation for use by other communities. However, the same knowledge might also need to be represented in a form that will serve the needs of system builders developing technology for the focus community. A common risk in planning a knowledge acquisition effort is to lose track of the intended audience for a particular KA workproduct, or to assume that a single representation will be able to serve multiple needs for multiple audiences. In Section 4.0 we will see how to combine these modes of knowledge acquisition in a plan for a realistic, comprehensive knowledge acquisition effort.

3.2 The Knowledge Acquisition "Canvas"

The name Canvas comes from a weaving metaphor; each element of the knowledge acquisition process — each investigator, each informant, each artifact — changes throughout the KA process. Thus the lifecycle of each of these elements is like a thread, weaving through the overall KA process. Each KA session, be it an interaction between an investigator and one or more informants, or a study of an artifact, represents the meeting of several threads. Each one changes a bit as a result of the interaction, and goes along to meet with other threads in other sessions. Viewed as a whole, the enterprise is a vast interconnected fabric of these threads. Planning for a KA enterprise does not mean planning all the sessions and all the paths for all the threads; since the threads include the learning lifecycle of human beings, we cannot pretend to be able to predict in advance how they will all develop. Instead, planning a KA enterprise involves preparing the infrastructure to track the progress of the threads through their lifecycle; determining what changes need to be kept track of, and what to expect of a session in which several threads are brought together.

Canvas is intended to invoke this metaphor of a woven network of threads guided by the Canvas framework. The knowledge acquisition planning process shapes the "Canvas," as it were, upon which informants and investigators will collaboratively weave the tapestry of acquired knowledge. Not all knowledge acquisition efforts will require a process that takes all the elements into account. The framework allows these choices to be made more deliberately, providing a basis for weighing various risks and potential benefits of different approaches. Consider Canvas, therefore, not a prescriptive recipe but rather a checklist of critical issues to be considered in planning and managing a KA enterprise.

3.2.1 Basic Elements

In this sub-section we introduce the basic elements or "building blocks" of the KA process in more detail: settings, investigators, informants, artifacts, and topics. In the broadest possible terms, information can be gathered in three ways: from artifacts, informants, or direct observation of work processes within settings. Information is gathered with a specific focus, centered around one or more topics. Information is gathered about particular work settings; artifacts studied are typically workproducts created and/or used in the settings of interest, and informants are practitioners with experience in current or previous roles in these settings.

In the subsequent sub-section we will explore how these basic elements weave together at various structural levels in specific KA sessions, threads that link multiple sessions for planning purposes, and the KA enterprise as a whole.

Settings

In the definitions, we introduced the notion of work settings and distinguished between the settings of the focus, investigator and target communities. These distinctions, however, reflect functional roles and relationships relative to the KA process itself, rather than the work practices under investigation. In order to develop specific practical guidelines for KA planning, it is helpful to explore in more detail the different kinds of settings that might be the subject of investigation using KA techniques.

The notion of setting implies, but does not require, physical and temporal co-location. A typical setting in the medical domain would be, for example, a doctor's office, night shift at an emergency room, or an accident scene for trauma care. Many KA techniques, such as the elicitation of scenarios as sequences of related tasks, are best suited for describing these kinds of settings. To describe an email discussion of a difficult case history among doctors, spread over a period of several weeks, as a "setting" is less intuitive.

Within a setting during a given performance period, many activities may take place involving many actors or agents. We assume that there are certain people who, typically because of job-related roles, routinely perform similar activities repeatedly within the same or similar settings. As participants in the community of practice in that setting, we refer to these people as ***practitioners***; practitioners in the focus community are called ***focus practitioners***. We distinguish this repeated activity, subject to learning, increased competency and expert status, as ***practice***. In the medical setting, patients are actors or agents but would not generally be called practitioners.

Investigators

Practitioners in the investigator community are called simply ***investigators***; their routine practice includes activities such as interviewing, cataloging results, and studying documents or tapes. The name "investigator" shares with its use in detective stories the skills of digging out information, determining what needs to be asked next, etc. However, KA investigators, unlike fictional detectives, are interested in more than information about individual events or persons. A KA investigator is also looking for information that is hidden within a cultural context behind unarticulated assumptions. This means that the KA investigator must also have skills of detecting the possibility of cultural bias in statements and awareness of possible ambiguity or misunderstandings of terminology. Also unlike fictional detectives, KA investigators rarely have to resort to fisticuffs or be able to detect cyanide from its aroma.

Informants

In a knowledge acquisition context, we use the word ***informant*** to refer to any practitioner in the focus community who provides information to the project. In Canvas, we use this general term, rather than more specific terms such as "user" or "expert" to avoid making any tacit assumptions about the status, within the focus community, of the informant. When planning an interview, however, the word used to refer to such a person can have considerable impact on how they view the process. This will be treated in detail later, when we discuss general stakeholder issues in knowledge acquisition in Section 4.1.2.

It is possible for there to be several focus communities in a knowledge acquisition enterprise; informants should be taken from each of them. Informants should also be selected to include many different roles in each focus community; this means that informants will include administrative personnel, software developers, domain experts, secretaries, support personnel, etc.

Artifacts

Artifacts are workproducts created and used in a work setting of interest, or containing information relevant to that setting, that are selected for study as part of knowledge acquisition. Use of the term "artifact" highlights several aspects: the fact that a sampling of workproducts is made as part of knowledge acquisition (i.e., not all workproducts get treated as artifacts), and the way in which the investigator accesses and analyzes the artifact may be quite different from how the same material would be used as a workproduct by practitioners.

More importantly, the investigator needs to be concerned with uncovering the implicit context out of which the artifact was created (a kind of reverse engineering) in order to make confident use of the material as a knowledge source. This means there is a potential interpretation step in working with artifact data that differs from simple gathering and transfer of data. As an example, suppose an equipment maintenance checklist is studied and it is noted that the mechanic has starred certain entries and crossed out others. Some interpretation is required to know what the stars indicate; for example, areas where trouble has been frequent, areas where a breakdown would have dire consequences, or areas that inspectors are most likely to check up on procedures. Thus an implied association of artifact as "cultural artifact" is not actually far off the mark. The KA process requires attention to the ways in which work culture knowledge and meaning have become embedded in the artifact. The artifact is "workproduct studied in context."

We also distinguish artifacts as "raw data" or "prior art" existing within the work setting independent of the KA enterprise, from KA workproducts created as *part of* the acquisition process. This distinction can be quite subtle. For example, suppose an investigator creates a high-level system architecture in conversation with an applications expert and documents this as part of the acquired knowledge from the session. Had the system designers chosen to develop such a diagram themselves, it might have been identical to the one produced by the investigator. Nothing in the content or representation format clarifies its status as an artifact or a KA workproduct created by (or with) the investigator's intervention. Only process traceability allows the distinction to be preserved. But this distinction will turn out to be critical in maintaining systematic links back to knowledge sources, and in correctly interpreting the material.

Topics

In Canvas, the word ***topic*** refers specifically to something known to the focus community that will be the focus of attention in some KA session. Topics are usually aligned with the overall objectives for the KA effort; for example, if the KA is being performed to elicit technology requirements then topics would probably involve exploring processes or workproducts (e.g., manual

forms) that were good candidates for potential automation. If business process improvement were the main objective, topics might include "breakdowns in work" or "customer complaints."

In the domain engineering context, topics usually fall within the scope of the domain. At various points of discussion in this document, we may use the terms "topic" and "domain" somewhat interchangeably.

3.2.2 KA Sessions

A knowledge acquisition ***session*** (KA session) is an event (or set of events) where an investigator consults some knowledge source (a person, a document, an observed process), elicits some knowledge and codifies some of that knowledge into a KA workproduct. For convenience, we treat all these activities as part of the session, even if, for example, the investigator leaves the interview and writes up session notes that evening in her hotel room. It is the full cycle through access of the knowledge source, elicitation, and codification that bounds the session as a whole.

A knowledge acquisition session has objectives that are determined as part of the planning process. The objectives include the following:

- *Topics of focus.* The focus of the session might be directed in various ways: for example, exploring the various kinds of tasks a practitioner performs in a given setting; or tracing the sequence of a given task or procedure in detail. The topics are basically a scoping technique that help direct attention to a tractable amount of material to cover in the given session.
- *Knowledge types.* The type of knowledge to be elicited can be used to focus how the session progresses; if the session will elicit procedural knowledge, then a different sequence of investigations will happen than if the session is intended to elicit declarative knowledge. Types of knowledge and the impact this can have on choice of representation notation are discussed in detail in Section 5.0
- *Audience.* While a knowledge acquisition project usually imposes an overall intended audience, each session can define its own more specific intended audience.

The topics, knowledge types and audience choices have some interdependencies, and will help determine the knowledge sources consulted for the session and the format of the session (e.g., one-on-one, joint meeting with multiple informants, walk-through of a document with an expert, etc.), and the representation used for the knowledge acquisition workproduct. For example, if the topic of focus is a task, then procedural knowledge is the type of knowledge required; certain representations will be more appropriate and certain types of informants will be better sources for this kind of data.

Each session will have a primary outcome that reflects some degree of having satisfied the objectives. Specifically, some knowledge in the topic areas and of the desired knowledge types should have been acquired and represented in a form appropriate for the intended audience.

There will also typically be some *unanticipated* knowledge acquired as part of each session. For example, suppose that during a session in which the investigator planned to elicit the steps taken by the informant in preparing a helicopter for flight, the informant provides information about the various ways in which the helicopter could malfunction. The investigator must be prepared for this situation, and decide whether to stick to the stated goal of the session, or to modify the session goal to pursue a new goal, say, of completing some part of the catalogue of helicopter malfunctions. The unanticipated information might be of relevance to the planning of the knowledge acquisition process itself, rather than the topics. For example, an informant might tell the investi-

gator about several other people that should be contacted as potential informants. This information then must be fed back to the planning process in some way; it may or may not affect the ongoing plan and schedule for knowledge acquisition.

In addition to the desired and unanticipated knowledge codified as a result of the session, each session has *peripheral* effects as well. The investigator leaves the session with a greater base of experience in the topic area and the work setting. The informant may have thought about his or her knowledge in a new way. Rather than treating these effects as spurious, or ignoring them completely, the Canvas approach to knowledge acquisition explicitly recognizes the inevitability of such effects. The peripheral knowledge gained during a session forms the basis of what we will call the "thread" of development of an investigator. A systematic knowledge acquisition process has mechanisms in place that support the management of both the intended objectives and results and the peripheral "side effects" of each session.

3.2.3 KA Threads

The Canvas framework reflects the essential idea that knowledge acquisition *creates new knowledge* through the elicitation and codification process. Each knowledge acquisition session creates knowledge both in ways that support the overall intention or goal of the session and in other peripheral ways as well. In particular, every session potentially has the following effects:

- changes the investigator;
- changes the informant; and
- (potentially) changes an artifact that is studied through the addition of annotations, commentary, or interpretation through the codification process.

This implies a sort of "Heisenberg Uncertainty" principle for knowledge acquisition. There is no such thing as passive knowledge acquisition; the process is always an *intervention* of some kind in the work settings of focus. To trace the impact of each session we can define separate learning "life cycles" for investigators, informants and artifacts. We refer to these cycles as ***threads*** The paragraphs below give a high-level overview of the major threads that weave through the various sessions conducted as part of a KA enterprise.

Investigator Threads

Since knowledge acquisition is a learning process as well as a transfer process, each session has two outcomes with respect to the investigator. As described in Section 3.2.2, there is the knowledge that was codified and made available to an audience community as well as knowledge that is not codified, which is nevertheless available to the investigator in later sessions. Exhibit 5 shows the development of an investigator's knowledge through participation in a series of sessions.

Issues faced in this life cycle include the following:

- Bias — information gained in an earlier session can interfere with information to be acquired in a later session.
- Preparation — information gained in an earlier session might be a pre-requisite for understanding information in a later session.

  Example. In the TCIMS effort, there were different types of sessions defined as part of the Knowledge Acquisition Plan. The TCIMS plan defined *baseline sessions* to provide a founda-

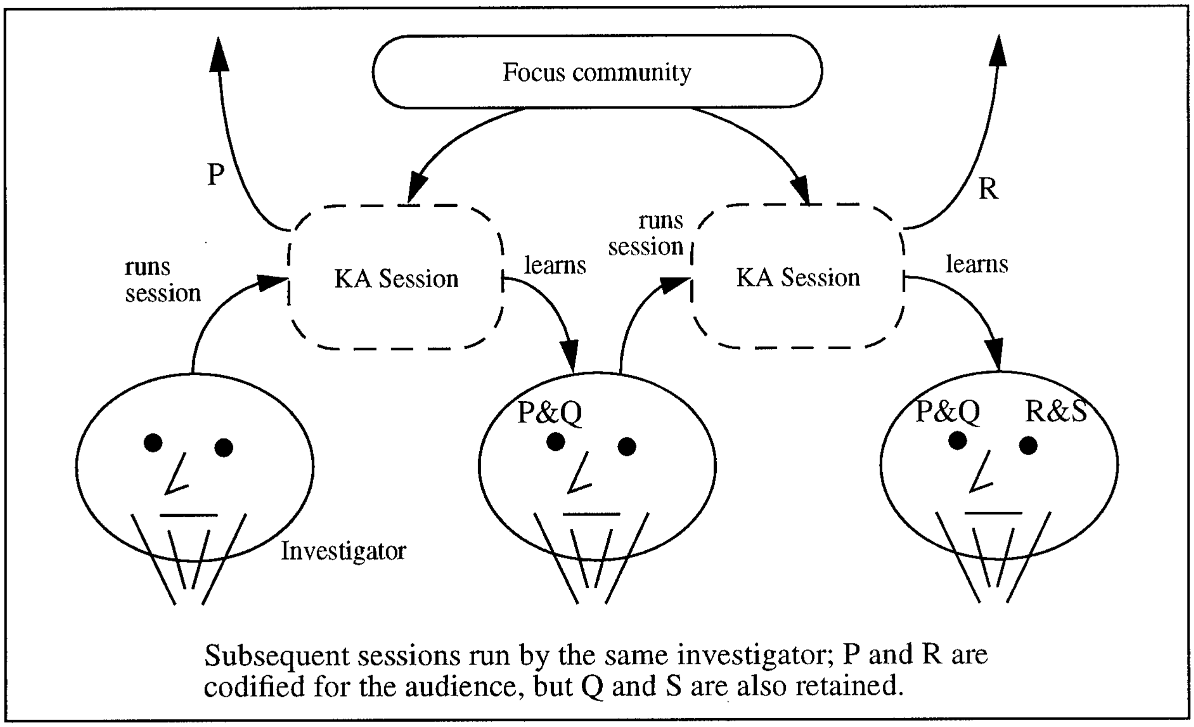

Exhibit 5. Thread of the Life Cycle of an Investigator

tion of domain-specific knowledge; they served both to orient investigators to the main technical areas and to orient *domain experts* (the preferred term for the informant role in TCIMS) to the knowledge acquisition enterprise itself. Later, more *specialized sessions* were conducted as part of the plan. For each investigator, a baseline session would ideally precede more detailed sessions in a given domain or technical area. In this sense, the TCIMS plan includes a mechanism for developing an investigator, by starting with a baseline session (either conducting it himself, or studying its resulting KA workproduct), and moving on to one or more specialized sessions.

In addition to the life cycle regulating each investigator within a given technical area, the degree of experience a given investigator has in general knowledge acquisition techniques, in the broader medical domain, or in the use of specific representations was a factor in selecting investigators for particular sessions. Over time, the plan could assist in the development of the investigator's skills in desired directions.

Informant Threads

Since the KA process creates new knowledge for the informant as well as the investigator, it is important to consider the sequence or series of interactions through which a practitioner becomes engaged as an informant to a KA enterprise (The informant thread notion is not sufficiently distinct from the investigator thread in terms of visual representation to warrant a separate illustration.). Through his interactions with the investigator, the informant also learns something, because of the intervention of the KA process on his own knowledge state. Therefore, the informant also goes through a learning lifecycle, which forms a thread of the canvas.

There are several critical issues to be considered in the informant thread. The thread creates increasing familiarity with the goals and approach of the KA enterprise. A certain amount of orientation is required to make someone, even an expert in the field, an effective informant. Many

informants will be interviewed only once; however, others may become important continuing resources for the KA project team. In considering the thread of multiple sessions with a given informant, therefore, planners should consider the role of the thread in gradually educating the informant to the knowledge acquisition task. This is almost the dual of the investigator's thread, which represents the investigator gradually gaining more detailed knowledge about the domain.

> Example. On the TCIMS effort, there were a series of specific representations and corresponding types of interviews defined in the plan: e.g., task analysis, scenarios, conceptual analysis, etc. Generally, there was a view that, for a particular informant, there was a most desirable sequence for these types of interviews: i.e., task analysis would be a more accessible starting point than conceptual analysis. This suggested that, for informants who would be significant knowledge sources interviewed several times, a certain "life cycle" of different types of sessions happening in a pre-defined sequence was an important aspect of the planning process.

There are also more pragmatic issues to be considered in the informant thread: e.g., making sure that the informant has been "tapped" for the important topics for which he or she has relevant knowledge, with minimum redundancy across sessions. Interviewing the informant in tandem with others from the same setting will also affect the KA process.

> Example. On the TCIMS effort, where medical experts' time was at a premium, an important task of the KA plan was to ensure that the same expert was not asked the same question multiple times. Instead, the knowledge acquisition reports were organized into a dossier that provided a structure so that all investigators could, in principle, access the data derived from previous sessions before scheduling and performing new interviews with a given expert.

Artifact Threads

Planning the life cycle for a specific artifact, as an inanimate information source, is a different problem than planning for the various interactions of an informant.

Clearly, an inanimate data source will not itself be affected by psychological phenomena such as bias or learning.[1] This does not imply that management of bias is insignificant in working with artifacts, which are strongly shaped by the embedded contextual knowledge of their creators and the work settings in which they are created. (In fact, in data gathered from artifacts rather than informants this hidden contextual information may be harder to detect and manage, and often can only be discovered through interactions with informants.) However, this initial embedded bias does not change as a result of subsequent interactions of investigators with the workproduct.

Instead, the thread for an artifact consists of the various phases of *interpretation* performed on it. In Exhibit 6, we see a sample life cycle for a knowledge acquisition artifact: each dashed line in the exhibit refers to the production of new KA workproducts, resulting from a study of the previous one.

The artifact begins its role in the KA effort as a workproduct created by practitioners within the work setting of the focus community. The workproduct is selected as an artifact for study (a knowledge source) by some investigator. Through a study of the artifact, new knowledge is elicited, which is then codified in a KA workproduct such as annotations of features of interest for the artifact. These annotations can be considered part of an ongoing "composite workproduct" that

---

[1] We exclude learning programs or other technologies that suggest the notion of an "intelligent artifact," though clearly these could be accommodated within the Canvas framework via a thread that allowed for the influence of previous sessions on the artifact itself.

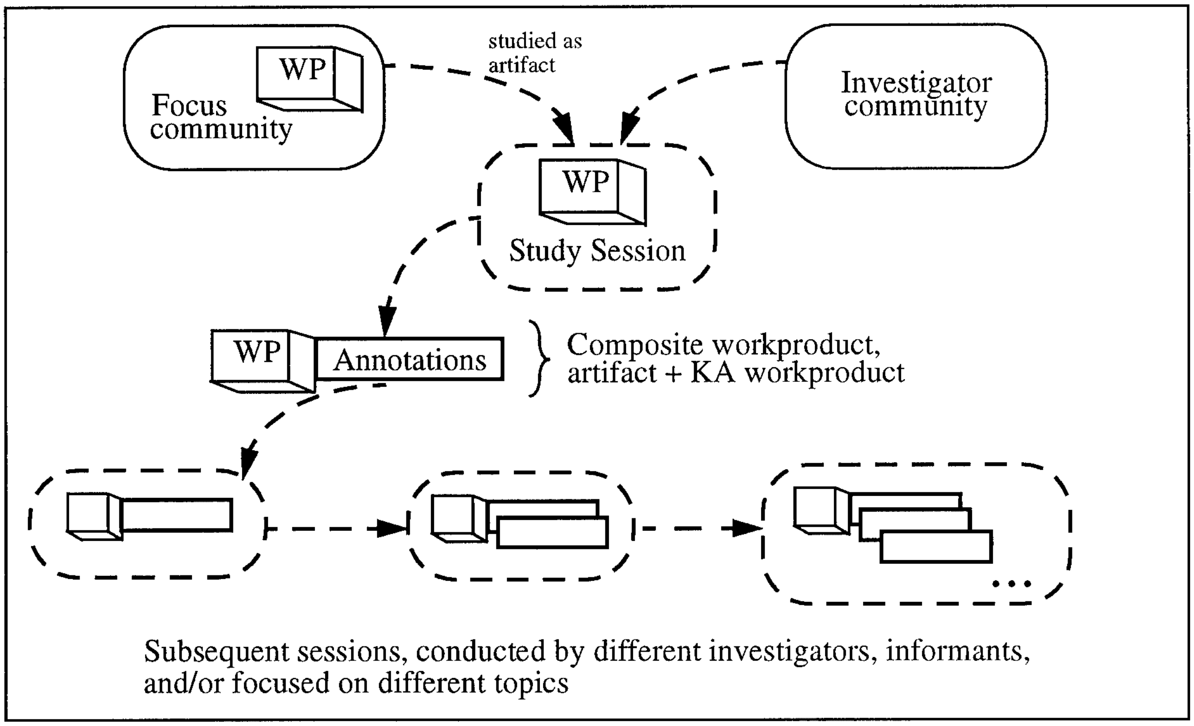

Exhibit 6. Thread of the Life Cycle of an Artifact

adds a layer of interpretation to the original workproduct. For example, subsequent investigators studying the same *original* workproduct as an artifact could choose to view the annotations of previous study sessions as well. There could be reasons for them to view these annotations or reasons for them to choose *not* to view them (e.g., controlling bias).

The composite workproduct (workproduct + annotation or interpretation) can itself becomes a subject for further study, perhaps by a different investigator. Through such sessions, the workproduct undergoes successive stages of interpretation; some common interpretations are annotation, comparison, and summarizing. This sequence of workproducts, beginning with the original workproduct from the focus community and continuing with successive interpretations added as part of the KA effort, forms the thread for that artifact in the Canvas framework. The KA workproducts in such a thread are called ***derivative workproducts*** (of the original focus domain workproduct).

Issues faced in this life cycle include the following:

- Redundancy — controlling multiple studies of the same artifact.
- Filtering — Determining how to filter the topical material from the artifact as a whole - particularly tough in domain engineering.
- Dropping through the cracks — Making sure that a given artifact has been examined, or that a decision has been made not to examine it.
- Scoping — Making sure that subsequent interpretations of the artifact fall within the scope of the KA enterprise.
- Proprietary access — Observing restrictions on who controls what can be seen, placed in dossier etc.

We show how these issues interact in the following example:

> Example. In the TCIMS effort, a number of reports were written based on initial interviews with informants. These reports are knowledge acquisition workproducts, and were stored in a repository called SEPWeb. Later in the effort, these reports were retrieved from the repository, and further analysis was done to them; these included summaries and changes of representation. The original reports were considered consortium confidential, and were not released for public viewing, while the derivative workproducts were published on the WWW. Further efforts summarized these workproducts further, to produce models written in a formal modeling language. Along its thread, we see changes in both the level of security and formality of the increasingly derivative workproducts.

3.2.4 The KA Enterprise

The KA enterprise is the project-level scope of planning that integrates the knowledge acquisition process into a larger context such as a technology development or domain engineering project. There are several aspects to the enterprise, including establishing the overall goals, making specific selection of resources (investigators, knowledge sources), and managing the ongoing process. These planning aspects are discussed in detail in Section 4.0.

An enterprise may address many topics, to be explored in many work settings. The enterprise may require a series of knowledge acquisition sessions to cover a given setting with great thoroughness or may rely on a representative sampling of data. As we describe in Section 3.3.3, if enterprise goals include systematic treatment of variability then a single topic may be investigated across multiple focus community settings.

The threads described above — investigator threads, informant threads, and artifact threads — are not independent. Each session potentially plays a role in several threads; i.e., either in the life cycle of an informant, investigator, or artifact. The sessions described separately in the thread discussions above could be the same session. In our canvas metaphor, each session is therefore a crossing of threads, a meeting point where each thread moves its subject forward along its respective life cycle. Planning a single session involves simultaneous changes to multiple threads. Conversely, each thread progresses by passing through a sequence of sessions, linked by the common element of the subject for that thread.

The essence of planning knowledge acquisition with Canvas is to manage development of all threads in parallel, according to the issues for each thread type as outlined above. This is a highly reactive planning task, one that requires careful attention to development of all resources of the project. The outputs of a given session that may affect subsequent decisions cannot be known in advance; planning must therefore be iterative and some effective means of communication of results across sessions, among investigators must be provided.

3.3 Issues Implied by the Framework

The basic definitions above provide a useful conceptual framework for a wide variety of KA applications. However, the emphasis of this guidebook is on two particular kinds of KA enterprises, typified by the applications to date of SEP and ODM. In this section we outline how these two applications influenced the design of the Canvas approach, and the ramifications of the framework.

The first primary application is the use of KA as an integral part of technology development, particularly of software engineering. Use of SEP on the TCIMS project is an example of such an

application. While KA is a well-accepted part of the expert systems development life cycle, it is by no means well recognized as essential to the system engineering process. SEP experience on TCIMS and related health care applications shows that KA techniques have an important contribution to make even when the systems proposed do not have a significant artificial intelligence technology component. Knowledge acquisition and the later knowledge modeling stages that translate KA results into more formal semantic representations can provide a basis for more effective systems requirements analysis. The same techniques can be used to suggest new potential areas for automated support in a work setting (i.e., pre-system concept development) and, at the back end of the process, can validate and evaluate use of a fielded system in its installed work setting.

The second primary application of interest in scoping this guidebook is use of KA techniques as part of domain engineering to support software reuse. In particular, the ODM process is typically applied to the task of modeling multiple software systems in a related domain in order to discover commonalities and well-scoped ranges of variability that can guide the design of reusable software components. While the SEP method has contributed a rich repertoire of specific KA techniques and representations, ODM has provided a broader conceptual framework to handle the range of KA tasks required for both KA-guided system engineering and domain engineering.

In extending the framework to cover both these target application contexts, we have of necessity generalized it to accommodate other applications as well. Traditional approaches to KA have limitations deriving from the evolution of these approaches from work in the social sciences or linguistics, or in knowledge engineering for expert system development. The necessary extensions include the following aspects:

- Knowledge acquisition is communication between cultures.
- Knowledge acquisition can be applied equally within technology-oriented settings as in work practice settings where technology does not play a key role. These applications, however, challenge certain assumptions in traditional KA terminology and approaches, require new techniques as well.
- Knowledge acquisition needs to support the explicit management of variability in the data.
- Knowledge acquisition, at its best, is a collaborative process, and
- Knowledge acquisition intervenes in the organization to which it is applied.

Each of these five aspects is discussed in a separate sub-section.

3.3.1 KA as Cultural Communication

According to the definitions given above, knowledge acquisition implies knowledge transfer across distinct settings; the cultural shift from one work setting to another is inherent in the process of knowledge acquisition, and cannot be avoided. Within each respective setting this knowledge may be expressed in varying terminology and put into practice in very different ways. This challenge could broadly be termed “cross-cultural communication.”

A well-known example of this challenge comes from the history of expert systems development, where knowledge embedded in the medical community had to be made available to a community of program developers and AI researchers in order to produce medical expert systems. Even within the medical community itself, there can be cultural differences. Most hospitals today are organized in such a way that there is considerable cultural difference between doctors and nurses.

Translation from one professional or work setting to another is thus fraught with the potential for cultural misunderstanding. Since we are often prisoners of our own cultural perspective, we are unaware that this cultural translation and re-interpretation process is taking place. In Canvas, we acknowledge this shift, and plan for it directly.

3.3.2 Technology-intensive Settings

Many methods for data or knowledge acquisition reflect a technology development context where the primary goal is automation of manual processes. Knowledge acquisition as part of social science research evolved from studies of non-industrialized cultures and were gradually applied to the culture of the researchers themselves. In the era when the first major computer-based systems were put in place, most business work environments were based on manually oriented work processes, with information exchanged via paper, face-to-face meetings, conversations, etc. Even early expert systems development tended to be initiated in domains where little automation was present because the decision processes were too intractable for contemporary software engineering techniques; thus the problems of integrating knowledge-based systems with legacy software systems only gradually came to the fore.

However, in current conditions a reasonable default assumption is that new or evolving systems will be put into place in a technology-rich landscape, with numerous existing legacy systems in various stages of implementation, usage or decay. Even opportunities for expert system development now occur more and more frequently at the frontier of established system capabilities. This creates particular challenges as well as opportunities for KA techniques, which must be adapted to shed light on how work practices evolve in relation to and are influenced by computer-based systems within the work setting. The challenges and opportunities of doing KA in a technology-intensive setting include the following:

- Technology is not created in a vacuum. It is created within a technologists' community of practice that has complex interactions with the user community where the technology is eventually fielded and used. If we want to study a technology-intensive environment and particularly pay attention to the technology aspects, we must gather data about both these communities.

  This is not a simple undertaking, however.Technologist and user settings generally represent quite distinct communities of practice. It is the norm rather than the exception that technology developers live in a different world than the technology users. These communities "collide" as it were in the work setting where the technology is used. Perhaps the most familiar example of this is software error messages; the categories of errors that are reported are decided and described by the developers who produce the software. When the users see these error messages, they are often inscrutable. This is not simply a problem of poorly describing the errors; it is a problem of what categories of errors it is important to distinguish for the two groups. A KA enterprise must have some systematic way of dealing with diverse communities and integrating the information gathered from each.

  The problem of "cross-cultural communication" arises in technology rich settings at least once, in the cultural clash between the developer and user communities.

- It can be harder to gather data about work practice mediated by technology using traditional techniques such as interviewing and observation. Since technology externalizes many routine tasks, expert users may not be able to explain their procedures out of the context of the technology itself. Significant processes may take place with the flick of a key at a keyboard and hence be nearly invisible even to direct observation.

  On the other hand, The presence of technology in the workplace also creates opportunities for

knowledge acquisition. For example, through instrumentation automated logging and collection of usage data can be performed in a relatively non-obtrusive way. Since every fielded system will need to evolve over time in response to needs and limitations discovered by users, some consideration for "system as knowledge acquisition aid" should be a routine part of the technology design process. It is also possible for investigators to experiment directly with legacy software as part of knowledge acquisition.

- If the KA is being performed to provide information to technology developers, a special set of stakeholder issues may arise around expectations about and resistance to the introduction of new technology. Even in situations where there are no clear requirements for new applications, technologists will have their own product ideas that exert influence on the sort of data deemed valuable and the kinds of questions that are asked. Similarly, users have ideas about desired system functionality, or resistance to introduction of technology, which affects the information they offer. KA project planning needs to take into account the influence of legacy systems in the setting and in the experience of practitioners, as well as the influence of trends, business requirements and expectations about changing system capabilities.

  However, this challenge is not side-stepped by neglecting the knowledge acquisition task as part of system development. Knowledge acquisition is deeply woven into the entire process of system specification and development; in fact, some form of knowledge acquisition must precede almost any requirements analysis phase. The kinds of issues dealt with in the KA enterprise should be of paramount concern to those creating technological systems for use in people's everyday work practice.

- It is tempting to model the environment of the prospective users of a system as a sort of "black box," which provides some input/output relationships that are exploited by the user. However, in technology-rich settings, it is often possible to change the environment's behavior almost as easily as it is to conceive the change. The placement of new technology into this setting will also modify the environment. This means that the users' environment must be modeled dynamically. Failure to do this will encourage "requirements creep," in which each addition to the technology causes the requirements on the technology to change. The knowledge acquisition effort must therefore pay special attention to modeling the relationship of the user community to its surroundings.

The ramifications of these aspects of technology rich settings will echo throughout the knowledge acquisition process. Knowledge sources taken from all relevant settings, including the informants as well as the developers of the systems they currently use or will use will have to be managed together in the evolving fabric of Canvas.

3.3.3 Systematic Treatment of Variability

Variability has not always been an element of explicit focus in conventional approaches to KA. It can refer simply to "noise" in the data. It can refer to lack of consensus in expert opinion. But explicit documentation of variability is important for certain KA goals; in particular, accommodation of variability is of prime importance in domain engineering. For the purposes of reuse, variability will generally surface in comparisons across similar systems within a specified domain.

Historically, knowledge-based or expert systems have been oriented towards domains of high-level, professional knowledge. Variability in opinion or belief may be interpreted as a sign of instability in such disciplines. Hence it is to be expected that informants might be reluctant to emphasize variability aspects, and/or that interviewers might simplify data to emphasize a common or generic result.

In previous applications of SEP, variability was dealt with primarily in the later, architectural aspects of system development. Variability was not addressed explicitly in the KA workproducts themselves. In the KA phase, investigators sought the common view of the experts they studied, which sometimes involved reduction of variability. How this common view emerges out of data with inherent variability is one of the points where the "magic happens" in a typical KA process.

In contrast, domain engineering for reuse, as exemplified by the ODM process, requires a specific search for patterns of variability. A specific "representative set" of knowledge sources must be selected that provides sufficient data on the variability to be modeled. The sample set of data selected and elicitation techniques employed are designed to explicitly *include* rather than reduce variability. In the reuse context, this variability serves as a kind of "second-order" data that establish criteria that help to define an intended multi-use scope of applicability for components to be developed. This presents certain unique problems for KA.

Although domain engineering highlights the need for systematic treatment of variability, the general concept of variability has rich implications for a KA framework applicable outside a formal domain engineering context. There are valid reasons to pay closer attention to variability issues in any KA effort. We can distinguish "various" kinds of variability that may need to be addressed in knowledge acquisition. Some are relevant to many KA tasks, others are more specific to a domain engineering context.

- *Variability within a given work practice setting*. Knowledge-intensive activity in any work setting typically involves making expert judgments among similar cases or conditions. This variability is the routine variability encountered in work practice; it needs to be negotiated by performers in that setting in order to perform competent or expert decision-making and judgment. An expert can recognize and classify cases, deal with ambiguity, and deal with "outliers" that somehow violate the assumed scope factored into the classification.

  Example. doctors in trauma care settings will encounter patients with a wide variety of trauma conditions to be treated, in conditions where there are insufficient resources (beds, supplies, medical staff available). Triage procedures in such an environment require being able to discriminate between a wide variety of specific injuries and conditions, and reduce these rapidly to a few categories relevant for immediate treatment decisions (e.g., lost cause, can survive without immediate attention, or immediate attention could make a difference).

  This kind of variability can be represented with semantic models (concept models, taxonomies) that capture expert practitioner knowledge.

- *Variability across work settings*. When multiple work settings are studied, variability can be observed across these settings. These variations may reflect differences in language or terminology, or variability in practices and procedures. Many representation notations for describing activities within a single setting provide ways of capturing alternative decision paths, etc., but not of reflecting diverse settings. These differences also might not be available as an established "model" from some informant, but might emerge as a composite of gathering analogous data in multiple settings, through a collaboration between the investigators and informants.

  Example. Continuing the triage example, triage practices might differ from doctor to doctor; there may be different policies in place at different medical institutions or in different countries, etc. These differences could provide valuable data in knowledge acquisition. Certain informants who had the benefit of experience in multiple settings might have greater awareness of these differences and be able to comment on them. Changes in practice over time within a given setting may also be a source of data, as observed by people with longer history of work practice within the setting.

- *Variability in system requirements.* An implemented system always reflects some implicit or explicit model of the work practice in the usage setting available to the developer. This model will accommodate certain dimensions of variability, in which case the system will be designed to respond to this variability. Conventional systems development, targeted towards building systems to support work in a given setting, has focused more on automating common, routine procedures, leaving performers free to focus more on knowledge-intensive tasks. Thus variability representations in conventional systems design focus on dynamic "procedural" models such as iteration, selection of alternatives, partitioning into process and data, etc. Expert systems might attempt to capture more of the variability that is a factor in expert performance. In any case, once the system is implemented, it provides a different source of data than the practitioner community.

  Example. Suppose that a computer system has been implemented to log the intake record and medical case histories of trauma care victims in emergency situations. Because of issues of liability it is desirable to record more than just the treatment administered to individual patients, since viewed in isolation decisions made following a triage protocol might appear to have been improper treatment. The computer system will represent an "abstraction" of the triage process that might embed assumptions of the process, the various types of rationale used in decisions, etc. The application might provide a hard-coded checklist of diagnosis codes, a checklist tailorable as part of system customization per installed site, a checklist that can be arbitrarily extended by the user, or even a free text field. This design decision, which yields a system feature in the implemented system, reflects a "theory" held by developers about where variabilities in practice might be introduced (i.e., different diagnosis codes), the range of that variability, and its desirability. New diagnosis codes could be entered over time, whereas variability in the form of misspellings or inconsistent usage of diagnosis codes would be counterproductive.

- *Variability across systems.* Domain engineering for the purposes of designing reusable software components must deal with variability at several levels simultaneously. The end goal is to create components or other assets that will be utilized by developers building software systems for a given domain. A domain-specific language is formalized to describe the semantics of particular components in terms of the domain-specific functionality they provide. However, since these components are intended to be useful in building many systems, this language must be expressive enough to cover the anticipated variability across those intended target systems. Each potential "application context," a system development effort in which the components could be used, will have been built (or will be built) with a specific usage setting in mind. The domain model must be able to express variability in those multiple usage settings to the extent that they affect the components themselves.

  This variability can be studied in at least two complementary ways. The first is through comparative analysis of multiple usage settings, or multiple system artifacts created for those settings. The second is by studying the development setting directly, particularly the processes performed by system developers building applications in the domain. It may also be necessary to collect information on a given system from the standpoint of both the processes by which it is created and the system products themselves.

Elicitation of variability information can pose particular challenges in the planning of a knowledge acquisition enterprise. The implications of a concern with systematic management of variability for planning the KA process is discussed further in Section 4.3.4.

3.3.4 Collaborative Knowledge Acquisition

A careful reading of the definition of knowledge acquisition in 3.1.1 reveals that there is no condition that the knowledge that is elicited, codified or transferred to a new community be "held" in any way by the informant. This is an intentional omission, which has considerable impact on how we view the knowledge acquisition process. In particular, although knowledge acquisition can include a simple transfer of known information from informant to investigator, the most profitable cases of knowledge acquisition are those in which some new knowledge is created as a result of the collaboration between investigator and informant.

A common complaint leveled against the field of Expert Systems development is that many of the systems are used only as tutorial systems to train new experts. The original motivation for most expert systems projects was that in some fields, certain knowledge was held by a small number of highly specialized experts, who would someday change jobs or retire; the organizations for whom these individuals worked wanted to capture this expertise as an asset that would last beyond the tenure of the individuals themselves. The expertise had been gained through many years of experience, usually with machinery or systems that were not well understood at the time, and hence was open-ended and full of mysterious items shrouded in the history of the system. No training materials were on hand to transfer this expertise to new practitioners by training, and the cost of apprenticeship was difficult to assess. The process of constructing the expert system included a phase of knowledge acquisition, in which this knowledge was organized, regularities were noticed and documented, and the resulting body of knowledge was then codified, often in the form of a rule-based system. This process of organizing and documenting the expertise removed much of the motivation for having an expert system; in particular, the expertise was no longer open-ended and mysterious, and it was now reasonable to imagine teaching it to a novice, or even starting a conventional software development project to support it. The rule-based system, which was intended to perform the expert's task in his eventual absence, was deemed a failure (since it never took over anyone's work). Instead, it took over the role of training material for new practitioners.

This process, by which the interaction between the investigator and the informant creates useful knowledge that was not available to either of them alone, is called ***collaboration***. The Canvas framework has been designed to treat collaboration as the default mode of knowledge acquisition, rather than as an unanticipated side-effect. This view challenges a number of assumptions about knowledge that might otherwise seem self-evident. We list a few of these assumptions, along with the contrasting point of view taken by Canvas.

All knowledge can be codified and transferred.

A background of formal education tends to make us think about knowledge as some sort of "material" that can be put into a book, and gained by reading it. There is something comforting about being able to hold up a book and say, "I know what is in this book," and being able to write down "everything I know" about some subject. Everything we know should be codifiable in some way in the printed word.

An extreme contrast to this viewpoint is the belief that no knowledge can ever be codified, and that anything that one really learns comes from experience. The Canvas approach takes a middle ground; accepting the fact that the result of any learning experience can only be partially codified, Canvas nevertheless relies on the assumption that *something* can be codified, so that it can be recovered by someone else at a later date. When that someone else comes from a different work setting, as is the case in knowledge acquisition, then the problem is to write something that can bridge the gap.

Knowledge is held by a single person.

A simple view of knowledge is that it is something that is known to a single person, and during knowledge acquisition, we will "mine" this person for the knowledge, which we will store somewhere. An interview is simply a transfer process. This view of knowledge is consistent with many interview situations, including most television interviews, where someone is debriefed to provide information that he alone holds. This view is encouraged by the metaphor of a "knowledge acquisition bottleneck," where it seems that there is some finite stuff called knowledge that has to be poured through a limited channel.

In contrast, in the Canvas view, knowledge comes from some interaction, either within the work setting of the informant itself, or in the interaction with the investigator. There is, of course, something very special about the informant who is chosen for a knowledge acquisition project; we could not take any person off the street, to use as an informant for the domain we are exploring. However, the result of the collaborative knowledge elicitation session will include things that were not known to that individual at the start of the session. These things might be beliefs internalized in the informant's work culture, or might be previously unremarked correlations among simpler facts of which the informant was already aware. The Canvas view on this issue is very similar to that espoused in [55], with the exception that Canvas examines its implications for knowledge acquisition. In particular, because of the variety of ways in which knowledge can be embedded in a culture, Canvas includes a variety of elicitation settings (group sessions, walk-throughs, reviews, solo sessions, ethnographic interviews, etc.) to reach this knowledge.

Practitioners don't use models.

In knowledge acquisition, some practitioner from the source setting is selected and used as an informant about some topic in that setting. Knowledge representation typically involves some sort of modeling of this practitioner's knowledge. A simple intuition says that in his own work practice, the practitioner does not use models of his own behavior; he simply acts.

In contrast, the Canvas approach is based on the recognition that while there are certain aspects of an informant's work practice that are inaccessible to description and introspection, many informants nevertheless do construct models of their behavior. In some cases, like professional medical practice, these models are quite thoroughly worked out and are very sophisticated. Canvas takes the view that there are different types of models for different purposes; a careful choice of model can exploit the informant's own modeling experience to encourage collaborative creation of knowledge.

Investigators reflect an informant's knowledge back to him.

Related to the mining view of knowledge acquisition is the view that an investigator acts as a mirror that reflects the informant's thoughts. The informant is not reflective in the sense of thinking about what he does; he simply acts and recounts, and the investigator does all the reflection. This is also related to the idea that informants do not make use of models. This view incorporates two assumptions; namely, that the informant is unreflective, and that the investigator is "neutral."

In contrast, in Canvas, the reflective expert is acknowledged and valued, and the role of the investigator is more collaborative. He participates in the reflection process of the informant by providing advice about representation, organizing principles, models, or whatever is needed to organize the investigator's reflective process.

Investigation can occur without intervention.

The idea that one can study human behavior without interfering in it has attracted a number of supporters from widely varying fields. Anthropologists have attempted to perform field work that does not intervene in the culture they are studying, while logical positivists view an expert as a set of rules, which can be taken from their context and used independently.

In contrast, Canvas is based on the observation that the most profitable knowledge acquisition will happen in circumstances in which the knowledge acquisition *does* have an impact on the informant and his work practice. The new, collaboratively created organization of knowledge will change how the informant thinks, and how he responds in new knowledge acquisition sessions.

A number of Canvas features provide specific support for this view of collaborative creation of knowledge, from the inclusion in the planning process of a thread specifically for tracking and managing the informant lifecycle, to the treatment of representations as aids for knowledge acquisition.

3.3.5 KA as Organizational Intervention

The idea of research as an intervention is not common, nor is it new. KA is a form of applied research that has exceptional capacity to appropriately intervene in a number of ways in the organization and management of work and workplace culture. From the Canvas perspective, such "interventions" always happen whether by intention or not. The Canvas approach embraces these changes only to the degree that they are bounded by ethical codes and are as focussed and as intentional as possible.

KA practitioners need to recognize that the research process itself can influence and intervene into the culture of study. As outlined above, merely making knowledge and practice explicit or asking a new question can change how an informant views his workplace. Such intervention can be positive or negative, depending on the purpose. The key is self-awareness and intentionality of the researcher and goal clarity of the research effort. Being too invasive by imposing unexamined cultural biases is always to be avoided.

A wonderful irony regarding this point is the following. "Pure" ethnographers make it a point of pride and ethics not to intervene in the culture under study. Yet, to some extent they always do, usually in unseen ways. Consultants and managers, on the other hand, are forever looking for new and better ways to intervene and change work cultures; yet they are constantly frustrated by the difficulties of making change. So, the one trying not to make changes can perhaps stir up more change than the one struggling to get change to happen. Go figure!

The KA process can serve as an intentional organization intervention in a number of ways, as follows:

Facilitate new conversations.

The LIBRA approach to assessment for reuse [45] stresses that organizations change by engaging in new conversations — new topics, new players, new ways of engaging players. The KA effort can bring people together who never knew about each other or never talked to each other or who held a simplistic or even contemptuous view of each other and their work. So it can serve to build bridges even with a given setting around the planning of the project, selection of informants, etc. KA sessions themselves are new conversations, and the resulting report or model can be used within the target or focus communities for enabling still more new conversations. The value of

new conversations is that they create shared language, understanding, motivation, and possibly ground for future change.

Hold up the mirror about work practices and contextual influences.

The KA effort focuses attention with the combined expertise of the investigator, the informant, and the connection between them provided by the organization and its management. The result is a greater capacity to examine what the work practice and the knowledge being used actually is. This can lead to clearer understanding of the need for change as well as the need to stay the same.

It can also show in dramatic relief how such organization contextual factors as structure, reward systems, training, job posting policies, health policies, and business strategy stifle or enable the day-to-day activities within a given practice community. Although these concerns are sometimes well below the view of certain practitioners, they can be of immeasurable value to leaders and policy makers who are truly unaware of (or who insist upon not recognizing) the impact of global decisions on local action.

Bring attention to work processes, customers, and collegial relations.

In the 1990s, attention to organization structure and process has shifted significantly away from vertical and function units that do not directly serve customers to lateral, cross-functional business processes that are customer focused. In a work environment where there is increasing pressure to think about work and management by lateral process, the work process discoveries available via KA are indispensable. Codifying knowledge in action is a significant activity that simultaneously pictures the work process as a series of steps or deliberations in the service of some customer value by a network of practitioners in collegial relation to each other. In short, KA efforts may serve to model these three key elements of organizational and process redesign via the knowledge codification process.

Bridge different work cultures, especially technologists and users.

A persistent problem in software-intensive user communities and software development communities is their difficulty in communication across the boundaries of their different languages, goals and customer pressures. Via collaborative modeling in the KA effort to discover new knowledge both in KA practitioners' collaborations with technologists and separately with system users, more grounded and objective ways of describing each community is possible. Where the possibility of greater understanding and connection between user and technology cultures is possible as a result of the outcomes of the KA effort, even greater synergy could be achieved.

A conscious KA effort to bridge the two communities might take a further step than collaborative KA modeling between informant and investigator. It might facilitate the collaborative modeling of knowledge that spans both settings by arranging interviews or focus groups so that the user and technologist are essentially doing KA with each other under the guidance of the outside investigator. Clearly, this is a form of KA which is also a facilitate intervention among stakeholders that uses the knowledge creation modeling and codification process to shift major entrenched elements of workplace culture.

4.0 Planning the KA Enterprise

This section builds on core concepts presented in Section 3.0 to present more specific guidance on planning a knowledge acquisition enterprise. This section discusses major decisions and issues to consider at various levels of planning, i.e., at the level of the KA enterprise as a whole, specific "threads" of the KA canvas, or individual KA sessions.

One of the major concerns that drives the planning process is the management of bias in the KA effort. Why must we concentrate so much effort on questions of bias in knowledge acquisition? The motivation lies in a key insight: The knowledge (or lack of knowledge) an investigator brings to a session is a two-edged sword.

An investigator's knowledge of a domain can enable him to ask good probing questions, to recognize accurate data and be cautious of questionable data, and can establish credibility with informants. Certain informants might be insulted to be interviewed by someone with insufficient background, particularly if the session is reduced thereby to a kind of tutorial on material available from other sources. This is a case where the informant, justifiably, would want to know that there was a qualitative distinction between knowledge transfer that primarily benefits the investigator and true knowledge acquisition that will render the knowledge more codified and accessible to a broader community. In the case of artifacts, certain artifacts might well be incomprehensible to an investigator without appropriate background.

Yet investigators' knowledge of the domain could also be a barrier in certain circumstances. For example, it can tempt investigators to "be their own informants" and impose opinions that reflect their domain knowledge on their informants, or otherwise bias the write-up of the acquisition session. Without formal validation of resulting KA workproduct(s) by the original informant the possibilities for such confusion become even stronger. Even for an investigator who is scrupulous about keeping his own views separate and partitioned from the data derived from his informant, the influence of previously held knowledge may affect what he does and does not ask and the terminology he uses in writing up the session.

This section is divided into three parts, one each for the three levels of KA planning, the enterprise, the thread and the session. Section 4.1, enterprise planning, discusses decisions that are made once for the entire knowledge acquisition project, while Section 4.2, thread planning, discusses decisions that are made once for the lifecycle of each element (investigator, informant, and artifact). Finally, Section 4.3 discusses the issues that remain to be planned for each session. Guidelines for planning to control bias and to account for variability are given throughout, as they apply to each level.

4.1 Enterprise-level Planning

For simplicity, we will consider that an effort called the ***knowledge acquisition enterprise*** or ***KA enterprise*** is to be initiated within a larger project context. In certain cases (e.g., social scientific research) this broader context might be almost coincident with the enterprise itself; but even in this case the proposed research is usually motivated (and financed) as part of a larger theoretical agenda, probably involving multiple researchers over a long period of time. In other cases, the KA enterprise will serve as an information-gathering phase intended to provide information to guide some intervention in the focus community. This scenario would correspond to KA's place within a broader SEP or ODM context, for example.

Isolated knowledge acquisition techniques can be applied much more informally within some project contexts. For example, a requirements analyst developing specifications for a new system

might interview some end-users about particular requirements or problems with the current system. A consultant preparing an assessment prior to recommending changes in business practices might do some qualitative information-gathering. While there is some flexibility in the definitions, we expect that the notion of a formal KA *enterprise* will not be particularly useful in these situations, even though they may meet the criteria for knowledge acquisition laid out in Section 3.0.

One critical element for a KA enterprise seems to be the need to manage global and local resources separately. A second element is that the various KA activities have numerous interdependencies, and reasonable independence from the other project activities. One interviewer may want to know the results of a prior interview in order to plan his activities; and many results might be gathered and collated before any hand-off to the target community takes place. This implies that scalability of the process might be another factor warranting a KA enterprise planning approach: i.e., the need to coordinate multiple investigators, informants, etc. Commitment to systematic management and tracking of knowledge sources, derivation of KA workproducts created, etc. is another defining element. Last but not least, in a true KA enterprise, perhaps as a result of some of the other factors above, the investigators emerge as a distinct community of their own. If the investigator considers himself to be entirely a member of the focus community or target community respectively, an essential "stakeholder" dynamic is missing from the picture.

In many cases, the investigator community will be formed specifically for the KA enterprise itself. There are a few cases where there is an ongoing community of practice that functions in the investigator role for repeated engagements or enterprises. For example, groups within large companies such as DEC or Xerox have been formed to provide specialized services of this kind as members of design teams for development of specific technologies. But we believe the norm will be that this community must be formed dynamically as part of the KA enterprise itself. Thus a central task of the planning process will be architecting the organization of the investigator community.

Note that typically the focus and target communities are *not* formed specifically for the KA enterprise. Unless there is a true community of practice involved for both focus and target, we are not dealing with the full set of issues involved in KA enterprise planning. For example, consider a journalist writing an article about a specific local culture (air traffic controllers, dog breeders, etc.) for the general public. In most respects this could constitute knowledge acquisition, particularly in that the journalist sets out to educate herself about dog breeders only inasmuch as this will help in writing the story. But in this case the "audience" for the KA workproduct is very vaguely defined in terms of the general mainstream culture in which both journalist and dog-breeders are situated. If the only way to define the target community of practice is "that set of people who may wind up reading my article" the Canvas framework is probably not appropriate.

Given the various provisos discussed above, the overall KA planning process can be divided into five areas of activity:

1) Setting objectives for the enterprise (discussed in Section 4.1.1);

2) Assessing stakeholder issues (discussed in Section 4.1.2);

3) Selecting various elements of the "canvas": investigators, informants, topics, settings, artifacts, etc. (discussed in Section 4.1.3);

4) Selecting the representations appropriate for the audiences of the KA enterprise (discussed in Section 4.1.4);

5) Putting an infrastructure in place to allow ongoing monitoring, capture of interim results,

cross-communication and coordination among different KA "threads" and to allow resource management (discussed in Section 4.1.5);

The last two points are of sufficient importance to the KA process that we have devoted full sections to them.

Once planning is complete and the infrastructure is in place, the KA activities proper can be initiated. As KA activities are conducted, the results are documented in the dossier, and objectives and plans for new KA sessions are iteratively adjusted in dynamic response to interim results. Detailed guidance in performing the KA tasks themselves (e.g., conducting an interview, observing work practice, studying a system artifact as data) is beyond the scope of this guidebook.

4.1.1 Setting Objectives

Planning any knowledge acquisition enterprise needs to begin with an assessment of the overall goals, in light of the larger project of which the KA effort is usually only a part. Important steps in establishing overall objectives for the KA enterprise should include answering these questions:

- What kinds of communities and work settings are the focus of the knowledge acquisition effort?
- Who are the various stakeholders? What roles will the various stakeholders (as individuals or groups) play in the KA enterprise?
- What KA "modes" best describe the overall configuration of the project? A large-scale KA project will typically combine several of the basic modes described in Section 3.1.3. This means that it is necessary to identify target and focus communities for the project. If informants are considered to be potential users of the information gathered, then the target community includes the focus community. If there is a third, distinct target community (e.g., technology developers) then the investigators will be playing a strong "translation" role between the communities.
- In light of these broad considerations, what are the specific objectives for KA, and how will they contribute to the overall project goals?

In addition to these general considerations for selecting objectives of the KA enterprise, special care needs to be taken to allow for the capture and documentation of variability. This will affect the choice of notations that will be used for representation of knowledge in the project. It will also have an impact on planning the various "threads" so that varying information will be available in a single setting, to allow for modeling of the variability. Some implications of handling variability in the KA process are:

- Potential barriers to capturing adequate models of variability should be noted as risks in the KA plan and strategies put in place to address these barriers. These obstacles might include the following: pressure on the part of practitioners to project a "normative" view of their work practice; difficulty on the part of investigators in representing the variability with adequate traceability back to distinct settings, informants, etc.; and unwillingness of technologists to work with data on variability that does not easily fit their model of what can easily be accommodated in implementations.
- Eliciting information for systematic documentation of variability may require specific techniques within the KA repertoire. These could include techniques for extracting variability data from workproducts such as interview reports or videos, where there was a different primary purpose of the data-gathering task. Viewing a number of artifacts or KA workproducts

in sequence, where there is a clear basis for comparison, is a quintessential technique for eliciting variability information. The clearest way to articulate variability is either side-by-side or successive viewing of multiple exemplars exhibiting common and variant features. This allows human pattern-matching cognitive skills to be applied. The knowledge acquisition plan might need to allow for collecting data from multiple settings, elicitation from multiple informants simultaneously, or analysis of similar or analogous artifacts "side by side."

Example. Suppose an investigator reads an interview report with a medical rescue helicopter pilot, then reads a second report with another pilot from a different facility. Provided that the purposes of the two interviews were similar and the performance of the data gathering session similar in nature (ideally, with similar questions asked) the investigator could begin to note variances and discrepancies in terminology, procedures described, etc. Here, it is precisely the placing of the separate workproducts in a common interpretive context that provides the basis for eliciting observations of common and variant aspects of the data.

4.1.2 Assessing Stakeholder Interests

In this section we consider stakeholder issues in KA planning. The knowledge acquisition process establishes certain canonical roles — investigator, informant, audience, at a minimum — that must be considered from the stakeholder point of view. Section 4.1.3 will describe criteria for selecting and characterizing these roles in terms of their potential value to the KA results. In this section we consider the three communities as stakeholders; i.e., what's in it for them? In particular, we are concerned with the following questions:

- What motivators and/or incentives do stakeholders have to participate in the KA enterprise?
- What possible barriers, perceived risks or disincentives might exist that present obstacles for their participation?
- What additional or synergistic outcomes could be defined for the KA process that might create added value from the perspective of one or more stakeholders in the process?

KA planning should involve consideration of each of these issues for each stakeholder in the KA enterprise. The temptation is, having identified some players with strong motivation, to stop the analysis and not go on to consider other players that may have strong interests threatened or compromised by the KA objectives. For each player, there may be conflicting incentives and disincentives. The planning process should involve building the complete strategic picture before attempting to develop piecemeal responses to particular issues and concerns. For this reason, we discuss below the common incentives and disincentives for all three communities of practice included in a KA effort, namely the focus, investigator and target communities. In discussing each of these stakeholder issues, examples are cited from experiences in applying SEP on the TCIMS project.

4.1.2.1 Focus Community

The focus community has characteristics that affect the structure and conduct of KA. These characteristics may be indicative of a subgroup of the focus community; further, the criteria for selecting informants from the population will be affected by these subgroups. In TCIMS, a body of experts in the medical domain were heavily relied upon for guidance and selection of high-value informants.

Motivators/Incentives

The KA process depends upon available artifacts and a pool of informants. Since informants must take time out of their work to participate in the KA process, it is important to consider what incentives are available to ensure that informants will be engaged and enthusiastic about participating in the KA process.

Key motivators for informants include: being treated respectfully; enhancing their status in their work community; participating in a project that will make a useful contribution; influencing key players they have been unable to influence; feeling heard, seen and recognized for having the expertise required to do the job well.

Treating the informant with respect, as in being on time, arranging a pleasant environment for the interview, and being personable may seem obvious, but can have a great impact on an informant's receptivity to the KA process.

Another motivating factor for informants is potential enhancement of their status in the focus community as a result of their participation in the KA process. Often, being chosen to participate in a KA effort may be felt to be of questionable value and even be treated humorously in thinking of the participant as the "victim." However, when KA investigators consciously respect informants and value their needs, other focus practitioners may actually want to participate. Hence, as participating becomes desirable, being involved can become a symbol of status enhancing the participant's leadership within the community and the value of her knowledge about the work.

Another incentive is believing that a successful result of the project served by the KA would make a useful contribution to the community of practice within which the informant is working. Informants' perceptions of the project will have a major impact on their enthusiasm and willingness to participate. Factors that will positively influence informants will be that the project is realistic (as opposed to a research "toy"), and that the project is feasible (instead of being so ambitious that success is unlikely). This motivator depends upon the investigator's ability to positively impress the informant about the value of the project in question, and the credibility of its anticipated results within the focus community. It may also require that investigators be able to explain why this project differs from other previous research or change projects; this may require in-depth listening to the informant's concerns about the project impact, as well as high awareness of previous projects that may have been or may appear to be similar to the current KA effort.

Similar to the above motivator, sometimes informants have been trying to influence colleagues, management, policies, etc. for some time and may be exhausted or have even given up. An astute informant may recognize the KA effort as a new and perhaps better vehicle for influencing practice, policy, structure, training, resource allocation, etc. This raises the problem of reverse confidentiality in that the investigator needs to be careful not to be treated as a beeline to the management for complaints. Negotiating a balance between being an appropriate channel for being heard and being treated as a messenger is important for the investigator to make this a benefit to both the informant and to the KA effort.

Similar to the above motivator, the frustration employees sometimes feel for not having their expertise seen and heard cannot be underestimated. The KA session has the potential for providing a forum for deeply listening, understanding, and appreciating employees' knowledge, expertise, and local, informal innovation. In many organizations, management holds a simplistic view of lower level jobs as being rote and mechanical. This simplistic view makes it easier to manage by the numbers or by objectives. According to the head of research at Xerox PARC, John Seely Brown, managing by this simplistic view is often less feasible than management would like[6]. He claims, regarding the informal daily innovation activities of even office clerks, "These infor-

mal activities remain mostly invisible since they do not fall within the normal, specified procedures that employees are expected to follow or managers expect to see." The KA effort becomes an ideal opportunity for employees to talk to somebody who really cares about these "invisible" activities and codify them as "knowledge" that may better teach managers new ways of observing and understanding their work. In short, investigators taking the time and gaining the knowledge in partnership with the informant may be of high value for some knowledgeable informants.

In general, the KA planner should consider ways of creating incentives for informant participation, including exploring synergistic goals that are peripheral to the main objectives of the project. Even if the KA materials to be gathered are primarily intended for use by technologists, the KA enterprise may create enough interest on the part of informants that they see ways to use the resulting dossier that were not anticipated in advance. The planning process should be flexible enough to explore ways to offer such spin-off value or secondary benefits to the focus community where possible (e.g., use of the dossier for training of new workers.)

Barriers/Disincentives

A number of factors may tend to discourage participation of potential informants, including simple though aggravating factors such as scheduling conflicts or total unavailability. In TCIMS, some informants were reassigned to the field in Somalia and Haiti just as their input was needed, resulting in restructuring of the informant pool. Expert informants are often in high demand, so scheduling may be non-trivial. Part of the rationale for careful management of dossier materials is to ensure that such experts' time is used as effectively and conservatively as possible.

In addition to these pragmatic issues, there are a number of other real or perceived risks that may act as barriers to informant participation. Informants will be naturally concerned about potential consequences of their participation in a KA effort. There may be distrust of the overall objectives of the project for which KA is being performed (e.g., as part of a "technology push" effort, or part of an undesired trend towards centralization and standardization of practice.)

There is a perceived risk in documenting de facto rather than regulation procedures. In many organizations, undocumented processes evolve to compensate for weaknesses in formal business structures and approved processes. Informants may believe that their description of these undocumented processes will cause problems with authority figures in the organization or will invite the scorn of professional peers. Thus informants may resist participating in the KA enterprise; or, if they do participate, may "fudge" or dissemble rather than describe the true situation as they see it. If the documentation of *de facto* procedures rather than or as well as regulation procedures in a given setting is an explicit goal of the KA enterprise, it is not enough to use techniques in the KA sessions themselves as a cross-check. The overall stakeholder picture must be addressed so that informants are not "tricked" into revealing data that may cause problems for them.

Strategies to mitigate these concerns must maintain the confidence of the informants. Responses could include anonymous aggregation of certain data, but this is generally more difficult with the highly qualitative data typical of KA than with sample quantitative data more typical of statistical studies or surveys. Sanitization of KA workproducts is also possible, but this risks introduction of systematic distortion in the data. Assessment techniques included in the LIBRA approach [45] involve indirectness such as gathering data via hypothetical storyboarding versus direct case studies. This can, in some cases, insulate informants from feared political consequences. In the case of TCIMS, where informants may have perceived such risks for themselves, the consortium agreed at the outset to keep all raw KA data as proprietary, thus guarding to some extent against repercussions to the informants.

Another potential disincentive for informants is the perception that documentation of their knowledge could compromise the job security or advantage offered by their own private expertise. This is certainly a factor in some settings, such as the threat letter sorters might have felt in interviews prior to the development of automated letter sorting equipment in the U.S. Post Service. In TCIMS, this did not seem to be an issue for medics, largely due to the performer-intensive nature of their specialty. To reiterate, medics have a common core of practice with an acknowledged high variability in how those core procedures are applied.

The issue above deals with expert informants who want to retain the advantage of their expertise within the community. The KA process can also create tension for informants who lack confidence about their own professional standing within the community. They may feel that participation in the KA process is "putting them on the spot" or even serving as a test of their competence in the field.

This potential barrier is one reason we have steered away from use of the term "domain expert" as a blanket term for a human knowledge source. For many kinds of KA tasks, we want to gather a rich set of data from different perspectives, including from the perspectives of practitioners in the work setting who would not consider themselves experts in the sense relevant within that community. This issue is particularly important when KA focuses on legacy systems being used in the field. Because of the nature of professional status in technology-intensive cultures, the people who interact most extensively with computer systems as users are often assigned relatively low professional status within their work settings. Yet their expertise in dealing with the current systems, getting information in and out of them, working around their limits, etc., may be highly relevant to the KA objectives for a technology development project.

Other strategies for mitigating this perceived risk include: giving informants adequate time to prepare for interviews; making the expectations and objectives of the sessions clear up front; making it easy for informants to defer to established sources within the interview for validation ("Check up in the manual on what I said; it's something like that anyway..."); keeping data confidential; and allowing informants to validate the write-up from their sessions to their satisfaction. There are, of course, associated issues with each of these strategies. At a minimum, KA planners must remain sensitive to the double-edged sword that the issue of "expert status" may present to their potential informants.

4.1.2.2 Investigator Community

Investigators have their own set of factors which motivate their participation in KA that need to be addressed. There are several issues that affect the investigator community, largely having to do with the organizational structure, project realities, and their own motivational factors. In this section we continue to draw upon TCIMS experience in the medical domain.

One significant complication for the investigators in TCIMS was related to project pragmatics and schedule. In theory, KA results should have provided input to the requirements specification phase of technology development. Given the iterative nature of the TCIMS project approach, the investigators on TCIMS were of necessity working *in parallel with* the technologists. These technologists were largely comprised of designers and developers of computer software and mobile computing systems. The parallelism of work meant that the investigators were continually adjusting their agendas to meet needs and expectations from both focus and target communities.

There was considerable time pressure placed on the TCIMS investigators. The parallelism of the project pressed them to create early results that could aid the requirements phase of the development effort. This created difficulties, especially considering that the effort needed to adequately

document sessions is considerable, being several hours per actual "contact hour." Schedules were further strained by the difficulty in getting access to informants.

Motivators/Incentives

In TCIMS, investigators were referred to as "KAs" or knowledge engineers. Because of their unique relationship with the informants (called "experts" in the TCIMS context) who were in many cases also decision makers, the KAs became some of the most effective ambassadors for the TCIMS project to the focus community (sometimes called the "stakeholder community" in the TCIMS context). This benefit of being on the front lines with the stakeholders was a source of considerable gratification to many of the KAs. Also significant to both KAs and informants was the incentive to see the KA process provide tangible direct benefits to the focus community. In cases where the KA team is serving as intermediary between domain practitioners (such as medical personnel) and technologists there is probably a natural tendency for KAs to adopt a bit of a "champion" role, to ensure that the interests of the focus community are being adequately addressed in technology development.

Investigators have a desire to see that the products of their labor intensive process are useful. The most tangible evidence of usefulness would be that the KA results were being used downstream in the project. In TCIMS, the time pressure and parallelism of the work tended to inhibit full use of the KA results in the downstream development efforts, which was a source of frustration for the investigators. In some cases there were prototype problems that could have been avoided had the information in the KA dossier been more fully exploited.

A source of both frustration and gratification for the investigators in TCIMS was that they were drawn more into the technologists' world than they desired. Some of the interactions between the KA personnel and the technologists amounted to their review of technologists' designs with respect to informants' expectations as the investigators understood them. This activity did cause the investigators to gain respect in technologists' community and resulted in the somewhat unexpected emergence of the KAs as the "ersatz users" or "answer persons" for the technologists, people who could provide information about what would really happen in the field, could partially make the bridge to the technologists' terminology and were more readily accessible than the experts in the field.

Barriers/Disincentives

The practice of knowledge acquisition is not a commonly accepted role in current technology projects, save those that are heavily influenced by performer experience, such as in artificial intelligence or similar technology. The result is that the KA is incorrectly considered "optional" in the hierarchy of project personnel, which is a major social disincentive to becoming a KA. The time and effort required to serve in the role of the "answer person" may also turn out to be a disincentive to some people. The investigator that prefers contact with the focus community may find contact with the technologists a distraction. Also, if a goal of the investigator or the organization the investigator serves is the development of transferable KA expertise then becoming the "answer person" may not be desirable.

It is worth noting that the above motivating factors are heavily dependent upon the investigator achieving a notable level of knowledge and understanding about the focus domain. This experience has motivated the design of the Canvas framework to explicitly model the additional domain knowledge acquired by the investigators in successive KA sessions as an important dynamic, with potentially both beneficial and undesired effects that must be considered in planning and managing the KA enterprise.

4.1.2.3 Target Community

The target community is affected by the KA process and by the community's interactions with both the investigators and decision makers in the various communities. We draw upon the following example from the TCIMS experience as a basis for the discussion, which is centered upon the technologists charged with building systems that operate in the focus community.

> Example. Technologists operating in the DARPA project environment experience conflicting expectations. There is a lot of pressure to construct demonstrations of varying fidelity and depth in a short time frame. The nature of the work ensures that there is an absence of consensus requirements to guide development. Further, in projects such as TCIMS, the consortium model of operation results in commercial companies each seeking viable product ideas, and this dynamic works against the competing dynamic of maintaining the aggregate team. The research nature of DARPA also tends to create expectations that participants will utilize other DARPA products, that often involve unproven technology.

Motivators/Incentives

In the TCIMS environment the technologists are in crucial need of usable and distilled domain knowledge. The ideal situation for the technologists would be if the final product of KA amounted to a real requirements specification. This is unlikely in practice. Other valuable input from the investigators would be tangible ideas for products. This has occurred at modest levels in TCIMS due to the investigators' interactions with articulate informants capable of describing desirable future situations. Another motivational factor for technologists is that the investigator may potentially relieve the technologists of the need to carry out a lot of time consuming basic research in the domain, provided information transfer from investigators to these technologists is efficient.

Barriers/Disincentives

Chief among the barriers to using the results of knowledge acquisition by the technologists is that the volume of written reports overwhelms their time and appetite for text. In TCIMS, this was a common complaint, although materials were readily available on a server. Three methods were used in TCIMS to mediate this problem; the KA person was used as an "answer person" in both a review and consultation capacity, a Web-based index of the KA results and derivative products was provided, and the large body of written reports were summarized and interpreted. This was partially successful, but the fact that the knowledge acquisition effort was being pursued in the same time frame as when the technologist were required to build and deliver the systems prevented realization of the full benefits of these strategies.

A partial explanation of the difficulties experienced with information transfer was that the TCIMS KA people were largely non-technologists. As such, they were not equipped to utilize formal notations. As a result, some less formal notations were constructed for their use that had the additional benefit of being more understandable by the informants. A desirable adjunct to the less formal notations would have been an automated interlingua that would facilitate the transformation of informal KA results to formal notation. This is a research topic in semantics that is not yet solved.

The technologists work under several conflicting constraints. First, they have the possibly conflicting goals based on being a DARPA project, as well as the commercial needs of their respective organizations. The KA results impose still further, often internally contradictory constraints. For example, in TCIMS, the complexity of requirements for systems that may be used by the complex focus community are overwhelming at the outset. Specifically, doctors, medics, and support personnel have varying requirements for platforms and interfaces.

Given that some of these constraints will have to be suspended or at least relaxed, it is tempting to apply pressure to the KA personnel to resolve some of these conflicts by subtly imposing a normative consolidated picture of the requirements, thereby biasing the interview process. From the point of view of the KA project, this risks skewing the picture of reality to the degree that it becomes a liability. This conflict will surface as a stakeholder issue from the target community side (i.e., there is some data they do not necessarily *want* to know.) For example, technology developers may not be naturally receptive to information that implies their technology or committed product ideas are not likely to be utilized in the focus community.

4.1.2.4 Other Stakeholder Issues

We have seen that the separate stakeholder interests of focus, investigator and target communities need to be considered as part of overall KA enterprise planning. Once these interests have been identified, the relationships of specific objectives with specific interests of stakeholder groups should be documented.

It can be helpful to identify potential areas of conflict up front. Not all objectives are compatible with the same project structure. As a simple example, if the KA enterprise is viewed by technologists as a kind of pre-requirements analysis to aid in technology design, and simultaneously viewed by informants as an opportunity to air concerns about institutional practices within the organization, these competing agendas could create tensions in the overall project and at the detailed level of the individual KA session (e.g., maintaining desired focus within an interview). The choice of representations for KA workproducts is one area where such conflicts are likely to become visible quickly. To document acquired knowledge in a way immediately useful to the focus community, amenable to validation by that community, or amenable to analysis by a target community may require very different processes and representations.

If the KA enterprise goals explicitly involve organizational intervention, this should also be recognized at the outset and appropriate change management issues addressed. The KA process may bring together people who do not interact as part of routine work practice. This may create new alliances and new perspectives on organizational relations, or may stir up potential conflicts. The KA enterprise planners should consider the possibility of these dynamics up front and have strategies at hand for responding appropriately.

4.1.3 Selecting the Elements

Once the overall objectives for the enterprise have been established and stakeholder interests have been assessed, the key elements of the "canvas" need to be selected. While these selections can be presented in a linear fashion, the various choices are highly inter-dependent. In general, the best selection approach is to first make the choices which are constrained by the project context and then select other elements to mitigate risks, address gaps and limitations or achieve synergies. For example, a project could be initiated where it is pre-determined that a certain group of practitioners will be the only knowledge sources directly accessible to the team. In this case, other knowledge sources (e.g., documentation, references, etc.) could be sought to create a cohesive overall set. In another case, it may be that the topics of interest are highly constrained but more strategic choices can be made with respect to the informants.

For each element type, a "pool" is selected as part of up-front planning. (The "pool" of investigators might more typically be called the "project team" or "KA team.") There are several advantages of selecting the pool as a separate planning step:

- Some selection criteria may apply to groups rather than individuals; for example, only medics from a given hospital setting will be interviewed for the project. The issues constraining

this choice (availability, stakeholder relations among organizations, logistics, resources etc.) will generally be at a higher level than the issues affecting choice of particular individuals to be in the pool. That is, the "pool" is more than a list of individual names.

- For each of the selection criteria for individual elements (described below), the pool serves as a high-level "reality check" that the overall project goals are attainable. For example, the KA objectives might require access to some informants with a high capacity for reflective thinking about their work practice. If it is determined that the qualities desired are rare within the community of access, this should be flagged as a high-level risk. It may mean the objectives should be down-scaled, or it may signal the risk that the project would be too highly dependent on a quality informant, if found.

Reference to a "KA plan" might imply an up-front, detailed plan that outlines how the entire project will unfold; but in fact, such a plan can provide only a starting structure. This is why we chose the name "Canvas" for our KA approach, to imply an image of weaving possibly new elements into the structure as the project moves along. After each session, new knowledge sources may be revealed that need to be considered in future planning. Sessions may yield more, or less, knowledge than anticipated; certain topics may be revealed as dead ends, while others, unknown at the start of the effort, move into the foreground. Just as a loom can be set after each woof strand to form a pattern, the KA project is set after each session, to take into account the effect it had on all threads. The planning activities that react to the changes in each thread caused by a session are properly part of thread or session planning; in enterprise planning, the pools of resources from which thread and session planning can draw are determined.

In this section, we give guidelines for selecting and characterizing the elements of the KA effort, starting with the work settings in which the focus and target communities work, followed by the three thread types, the investigators, informants and artifacts.

4.1.3.1 Selecting and Characterizing Settings

One of the most important decisions to be made in planning the KA enterprise is to determine the settings that are to be studied. This is important for several reasons:

- Informants taken from different settings will often use a different terminology, which might overlap with other terminology in misleading ways.
- Informants from different settings will have different levels of familiarity with various formalisms.
- Some source settings might also be target settings, in which case the informants can be treated as potential customers.

The settings identified in this activity can have a number of relationships to one another. In addition to the source/investigator/target relations described in Section 3.1.3, either the source or the target setting might consist of several culturally distinct groups. Some of these distinctions will be of interest to the data acquisition effort, and some will not.

> Example. Surgeons, medics, nurses and paramedics all draw on similar training for their terminology; therefore, as far as terminology is concerned, distinguishing between these three groups might not be considered relevant for a KA effort. However, among the on-site trauma personnel, there are also medical technicians, transport personnel (ambulance drivers, helicopter pilots) and other support personnel (local fire fighters, policemen). If any of these will be involved in the knowledge acquisition effort, then it is appropriate to identify them as separate communities of practice, based on differences in terminology. Among the medical per-

sonnel, the difference in professional status and responsibilities indicates that nurses and surgeons might be considered as separate communities, if the goals of the effort include information that is embedded in this distinction.

The determination of what groups should be considered as separate, and what groups should be used as sources of information, can depend on several characteristics of the work practice in the field. For each of the following characteristics, we provide some hints about the impact it can have on the knowledge acquisition process. Use these characteristics to determine the distinct settings.

Nature of expertise in the field

Is the field process-intensive or performer-intensive? In very mature fields, accepted practice is so formalized that it has been codified into a rather rigid process. A prime example of such a field is civil engineering, with a base of knowledge of practice several thousand years old, and where the cost of mistakes is such that process formalization exists in part to reduce risk. Additionally, in civil engineering, the behavior of the materials and forces that affect structures is relatively well understood, both due to hard-won experience and due to the development of analytical models that have some predictive power. This is not to say that performance in such fields does not have a subjective aspect. In civil engineering, the subjectivity is the artistic license that the engineer exercises in making an aesthetically pleasing design. The maturity of the field makes it possible to have a clear distinction between what is subjective and what is not. For the objective part, performance can be judged by how well the process is followed; legal liability may even be associated with failure to follow the process. Such fields are characterized as *process-intensive.*

Medicine, in contrast to civil engineering, could be characterized as *performer-intensive.* Although the practice of medicine has a similar long history, the repeatability of results is drastically less than in civil engineering, in no small part due to the complexity of human physiology, the variability of effects of treatment from individual to individual, and the changing cultural norms and accepted treatment practices. As a result, there is an accepted body of practice that leaves much discretion to the practitioner working within the guidelines of that accepted practice. In such a field, it is difficult to determine when a particular performance should be judged on objective grounds, and when it is part of the subjective skill of the practitioner. In medicine, the subjectivity lies in the professional discretion of the diagnostician.

Communities in process-intensive fields are more likely to be interested in an accepted, consensus view, while in performer-intensive fields, a careful study of variability is likely to be useful. Knowledge acquisition in process-intensive fields is more likely to encounter resistance based on fears on the part of the practitioners that their positions might be jeopardized, if the KA project is successful.

Nature of training in the field

The general approach to training in a given domain has a direct impact on the approaches one takes to deriving information from the informant pool. Disciplines in which analysis and process are highly prized usually offer informants who are trained to reflect upon their practice in detail and can offer well-structured accounts of their work. By contrast, trauma medics are trained to carry out procedures in response to recognized situations. The time critical nature of the trauma domain makes it very important for the practitioner to assess a situation, match the pattern of the situation to the procedures needed, and act with dispatch, else the well-being of the casualty or injury may be irrevocably compromised.

A similar distinction is reported by Brigitte Jordan as a result of fieldwork on the effectiveness of training programs for traditional birth attendants in the Yucatan [15] funded by industrialized nations. Her research suggests that the low effectiveness of much of this training results from misapplication of didactic modes of teaching in cultural situations where learning by an apprenticeship mode is more appropriate and culturally familiar. A training style based on Western norms of professional medical practice typically makes heavy use of sequential scenarios that move through the various steps involved in a medical procedure such as assisting at a birth. In contrast, an apprenticeship-based model might categorize tasks according to criteria such as the degree of skill, personal authority and experience required to carry out the tasks. Such a model would thus approach an overall scene description in concentric circles of peripheral support tasks (suitable for performance by a less experienced apprentice) and core tasks to be performed by the expert. Jordan was initially surprised to find the traditional birth attendants reluctant to share stories, which correspond to what is familiar in Western medicine as "case studies", a standard way for physicians to communicate both among themselves and with other communities. Hence even the ability to elicit knowledge from an informant via a sequential, hypothetical scenario already presumes a great deal of cultural context that might not be applicable in other domains or other cultural settings.

Stable versus fluid domain knowledge

The state of knowledge in a given field has a major impact upon the approach to utilizing informants or artifacts in the field. A domain might contain considerable amounts of stable knowledge. For example, in the civil engineering domain, there is a substantial body of settled knowledge relating to aggregate experience about constructing common structures within bounds of size and characteristics of local geology, given the parameters of the building materials. If some part of a domain is undergoing rapid change due to immaturity or high-impact advances in technology, the "half-life" of knowledge for new practitioners who deal with that part of the domain is shorter. In civil engineering, innovations in composite materials make the half-life of any knowledge based on the use of known traditional materials very short. Medical procedures in trauma care and emerging fields such as personal communications systems also have this fluid quality.

The overall objectives for the KA effort need to be appraised for realism relative to the rapidity of change in the field of interest. Domains that incorporate dynamic knowledge will require some more stringent techniques for cross-validation. Strategies used in the TCIMS project included combining interviewing with observation or cross-checking data from experts with knowledge of varying degrees of outdatedness.

Stage in business lifecycle

Businesses often go through a lifecycle that includes, roughly speaking, a start-up phase where the knowledge is volatile and innovative, a maturity phase where the knowledge is stable and well-accepted, and a decline phase, where the knowledge becomes obsolete and irrelevant. There is a trade-off of the effectiveness of KA against the need to do it in each of these phases. During the innovation stage, the need for a well-organized view of the knowledge in the field is high, since it allows technology planners to see the directions that the field can take. Unfortunately, it is also during this time that the value of the KA results are short-lived, and the time taken to complete a systematic KA process might outlive the usefulness of the knowledge. At the other extreme, when knowledge is stable in a mature business, the need for organizing the knowledge is less, since the field has managed to do some of this organization in its own practice. The effectiveness of KA is, however, high; since the knowledge is well-accepted and stable, it is possible to get a solid, coherent corpus of information about the work practice. During decline, the need for knowledge acquisition might rise again, as the number of working practitioners declines, and the knowledge is in danger of dying out.

A setting for knowledge acquisition should ideally be chosen from a business whose position in its lifecycle balances this trade-off, so that the KA is effective enough to provide value, and the need is high enough to appreciate this value.

Degree of professional stratification within community

Unless the focus community is rather uniform, such as professors of philosophy or long-haul truck drivers, some degree of professional stratification is likely to be observed within a focus community. The domain of medical practice is highly stratified, both in training and legal certification of the practitioners. Physicians have traditionally considered themselves to be the central focus of the medical field (although this is certainly affected by the emergence of HMOs.) At the same time the experience in TCIMS is that while physicians are usually the prime source of information on treatment and diagnosis, they have less detailed knowledge of the enterprise of medicine than other practitioners. A compelling example is that nurses are more knowledgeable about the actual workings of hospitals than physicians because of the division of labor that exists between nurses and physicians in most hospitals.

This state is not unique to medicine. It is common knowledge that the secretarial staff in most organizations has more immediate knowledge about the real workings of the businesses in which they function than the acknowledged decision makers. As a result it is suggested that the focus community be covered as completely as possible by the KA plan, including subgroups that might be considered as "non-professionals" to ensure that important features of the domain are not obscured or misrepresented by the KA products.

4.1.3.2 Selecting and Characterizing Investigators

Depending on the structure of the project, there might be a possibility for selecting the pool of investigators who will be performing the actual KA activities. Even in situations where the investigators have already been chosen, it is worthwhile to understand the capabilities and discriminating features of the investigators, to provide input to decisions about who will perform which sessions. The following skills required of an investigator are subtle and often contradictory; informants should be selected to cover these skills as needed:

- *Familiarity with the focus community.* A clear advantage in some situations for knowledge acquisition is if the investigator already has a familiarity with the field of study. This can be particularly crucial for artifact analysis, since it can be quite daunting to try to understand an advanced document without familiarity with the basic terminology of the field. This can also be a relevant factor in interviewing experts with very high demands on their time; if they need to spend time explaining fundamentals of the field, then they will lose confidence in the KA process, and feel that their time is not being well spent. Sometimes a familiarity with a closely allied community can reap some of these benefits. In the extreme case, a practitioner from the focus community itself can be recruited as an investigator.

- *Unfamiliarity with the focus community.* Familiarity with a field can bias an investigator in such a way that he will be blind to certain unarticulated knowledge that is embedded deep within a community of practice. Given that such embedded knowledge is often the most elusive goal of a KA project, the effort should include some investigators who are intentionally naive with respect to the focus community. Such investigators should, however, be aware of how embedded knowledge can manifest, so that they can take advantage of their naivete; otherwise, they are simply an underinformed investigator.

- *Ability to "play dumb".* In order to allow an informant to offer information in a form with which he is comfortable, it might be necessary for an investigator to allow the informant to

report information with which the investigator is already familiar. By acting as though he does not already know the things the informant is reporting, the investigator can encourage the informant to cover basic material thoroughly. This could well allow the session to uncover some normally unspoken assumptions that are embedded within the work practice.

- *Ability to "play smart"*. When an informant moves into territory that is new to the investigator, there is a good chance that a lot of knowledge can be acquired. If, however, the investigator seems to be getting lost, this will encourage the informant to retreat to firmer, simpler ground. An investigator who can give the impression of understanding all the intricacies of an explanation can encourage the informant to move into new areas.

- *Fluency with notations used for representation.* In a particular KA project, a large number of different notations might be used for the representations of the KA workproducts. Some of these representations will be intended for feedback to the informants, while others might be of use to system developers, and others might be for internal record keeping by the KA investigator team itself. Some of these notations can be difficult to use, and fluency with them is a skill that could discriminate one investigator from another. Some sample representations, and guidelines for their use, are given in Section 5.0.

- *Ability to bridge between communities.* The investigator often becomes a sort of ambassador from the focus community to the target community.This means that he should be able to understand motivations from both points of view, and be able to speak a language that is comprehensible to both communities as well. This ability also relies on the personal interaction skill of being able to switch hats, and fit in to either community.

From this set of skills, it would be surprising if any human ever gained the objectivity, creativity and receptiveness necessary to be an investigator in a knowledge acquisition project. But in fact, experience with other related activities can develop some of these skills, and the investigator can be effective even as he is developing the skills outlined above. Toward this end, we posit a knowledge acquisition skills "maturity model" that illustrates the various levels of skill that an investigator might already have reached by virtue of other, similar activities:

- The first "level" of knowledge acquisition skill recognizes that many of the skills of a good reporter are relevant to knowledge acquisition, such as building rapport, remembering what is said, taking effective notes, balancing leading and following in the interview, and taking care to cross-check and verify stories. Good reporters also have less tangible skills like "smelling a good story," being resourceful about finding people to talk to and other sources of information, etc. Ethical questions that arise for reporters such as balancing the impact of a story on the subjects and their community with the public's "right to know" can be understood in terms of conflicting stakeholder interests. Investigative journalism and detective work involve many of the same skills. Note that in classic journalism there is no imperative to report using the terms and concepts of the focus community; the reporter acts as the "eyes and ears" of the larger culture.

- The second "level" of knowledge acquisition skill can be differentiated by going beyond news reporting or detective work in that it attempts to elicit knowledge from information sources (informants) without imposing concepts, categories, or language from the reporters' or detectives' culture. This kind of concern would be more typical of a professional ethnographer (e.g., a cultural studies researcher) for whom capturing the native or emic categories is a specific goal.

- Unfortunately for the goals of knowledge acquisition, each session is inevitably an intervention in the focus community. The next "skill level" involves not merely seeking to have no influence, but acknowledging and attempting to plan and compensate for the influence that occurs. An inter-cultural reflectiveness is required to recognize what categories, values, and

language the investigator brings as a member of the community doing the research (in addition to her personal biases). This is particularly important because it not only colors the data acquired, but it does so in a way that is often very difficult to see in the data itself. To become aware of these factors means owning up to the politics of how "we" who are doing the studying may impose on "they" being studied. Sometimes this type of research reveals as much about "us" as it does about "them." In Canvas, the emphasis on looking at all the various stakeholder issues involved is one step towards fostering this skill level in KA.

- In what we view as the highest skill level in KA (within the scope of this document) the informant and investigator are engaged in true collaborative knowledge acquisition, i.e., they see themselves as working together to create a codified model of knowledge in the topic areas. True collaborative knowledge acquisition requires that investigators and informants engage in a relationship that (1) temporarily suspends the cultural lenses of their respective work settings; and (2) builds enough trust to discuss topics that may be culturally "taboo," politically sensitive or so deeply internalized as to be hidden to practitioners.

The description above places demands on someone who performs knowledge acquisition that may take years to master. Although we have presented this in terms of skill levels, the analogy to a true maturity model may be misleading here. Different contexts impose different needs for handling various issues (e.g. a reporter need not aspire to be an investigator in a KA effort.) It is possible to carry out knowledge acquisition tasks without having mastered all the subtleties of a collaborative, ethnographic approach.

KA Levels Needed for Software Technology Settings

A primary reference point for Canvas is use of KA in building better software systems; what level of maturity is needed to do this well? We believe that collaborative KA is an appropriate paradigm for this kind of KA situation. When software developers and end-users are both viewed as potential informants, the assumption that the informants cannot themselves do reflective, taxonomic or knowledge-creating activity breaks down. In addition, software technology artifacts (systems and the other workproducts used in creating them) lose their status as "received wisdom" and can be viewed themselves as cultural artifacts reflecting embedded views and assumptions and amenable to revision.

In some collaborative settings, the investigator provides the tools for representing the informant's knowledge in the form of a computer program. In this case, the expert constructs his own models with the help of the program. The Repertory Grid method [34] provides extensive model visualization tools, allowing the expert to input and manipulate his own models. An extreme version of this in the context of expert system construction is reported by Welbank [54], in which the expert learns to program a rule-based system, and maintains the rule models of the expertise himself. This can raise problems in practice if the size of the rule base grows beyond the expert's patience to maintain it, but there is no barrier in principle to having an expert learn how to construct even very sophisticated models.

One point we have tried to demonstrate, however, is that knowledge acquisition involves a coordinated set of cognitive and interpersonal skills that are as valid an area of specialty as technology expertise. Availability of such expertise can add value to a technology development project, but may need to be managed in distinct ways. There is nothing inherent in the training of technologists that make them particularly capable at this set of skills. Thus "drafting" technologists to play this role without special training or acknowledgment of the separate processes required may not yield the benefits of systematic KA. On the other hand, the requisite skills also challenge some of the assumptions of traditional knowledge acquisition researchers. Ideally, the Canvas investigator would approach KA as an opportunity to facilitate knowledge creation and sharing between tech-

nologists' and end-users' communities, potentially transforming the technology development process itself as a result.

Self-assessment profile

The following questions are a useful "starter set" for self-assessment of candidate knowledge acquisition personnel (investigators). Readers who aspire to be knowledge engineers can make use of this material to determine where they lie in the "skills level" profile above, and to assess what skills they need to develop to move forward.

- How much do I know about this topic area?
- How fast am I at picking up knowledge in new areas? (Quick study?)
- How facile am I at projecting an air of intelligence in areas that I do not know about? (Good bluffer?)
- When I am interacting around a topic I do know well, how good am I at suspending my own knowledge and opinions, in order to elicit the knowledge of the informant? (Imposer or good facilitator?)
- How attentive am I to reflecting back the terminology of informants in the interview? Do I introduce my own terms and expect them to shift? Do I check out and verify whether my usage of unfamiliar terminology is accurate?

There is fascinating interplay between these acquired skills (or traits) in knowledge acquisition. For example, does a person who can readily suspend their expertise in a familiar area also tend to be a person who can facilitate well in an unfamiliar area? We speculate that good listening skills are a fundamental requirement in both cases.

Skills required for collaborative knowledge acquisition

Collaborative knowledge acquisition expertise includes the skill of being a "skillful non-expert" who can help experts reflect on and articulate their own knowledge in fruitful ways. The investigator many times must act more as facilitator than an interrogator, by knowing the right questions to ask. The investigator must also, in the terms of Donald Schon, be a "reflective practitioner" [29] continually practicing the intensive inner work of examining and adjusting for personal and cultural bias. She must be ready to experience the ground shifting in the interview, discover and suspend successively finer details of this bias, without feeling defensive or, alternatively, fleeing her own cultural bias in favor of the bias of the informants. This last point is the dangerous loss of perspective called "going native." The investigator must also hold her own need for closure in suspension in the face of the emergence of data that is not immediately consistent, clear, or coherent. Maintaining high receptivity without judgment or premature conclusion in these situations is a difficult and a critical skill. The investigator builds a trusting relationship, but in a way that is not exactly mutual and does not allow the "interview" to lapse into normal friendly conversation, problem solving, advocating, or advising.

When the informant temporarily leaves the flow of his cultural daily life to reflect, talk and allow his knowledge to be elicited by the interviewer (who is suspending the lenses of his own culture described above), then a third, "collaborative" vantage point is established. This could be thought of as a temporary culture in which both informant and investigator can "play" as equals, confronting even harsh realities and envisioning new possibilities. This is where innovation may take place in a way visible to and acceptable to both communities of practice.

Investigators are selected and characterized based on their sets of skills and degree of maturity in applying those skills; this information will be useful in planning out the lifecycle of the investigator (her "thread") throughout the knowledge acquisition project.

4.1.3.3 Selecting and Characterizing Informants

In Canvas, we encourage the consideration of informants from a number of settings, and at varying levels of training or skill. This means that there is a wide array of considerations that can be taken into account in determining who to use as an informant.

Breadth of experience

An informant who has played several roles in the work setting, and has seen several changes in the work setting usually has a unique insight into how the setting works. This is not the same as depth of knowledge; even a casual user of many different processing system has quite a sophisticated idea of what they can and cannot do, and can be a great source of comparative data, even if the user has never learned enough about any of the systems to customize them.

Obsolescence of domain knowledge

While it may seem obvious that one should avoid selection of informants whose lack of current involvement makes their input suspect, this is not altogether adequate as a criterion. For example, there may be a trade-off in the availability of informants with the degree to which their knowledge is up-to-date. Also, depending on the stability of the knowledge in the field, being up-to-date might not always be the topmost concern.

Some care should be taken in complex domains to allow informants to validate their own input. For example, the intense field experience of knowledgeable combat medics in the U.S. military now dates from the Vietnam war era. In this case it is necessary that medics be given the opportunity to provide their recollections of that experience, review draft session reports and note where current practice diverges from that reported. Further, the aggregate session reports may be examined by the aforementioned body of experts in the domain for a more global validation. In this case, not only can experts with less current knowledge do some of their own validation, but new knowledge may be created by making the difference between older and more current practice more dramatically visible.

An extreme example along these lines is the experience of former air traffic controllers, brought back after lifting of the long-standing ban on rehiring, described in [11]. In this case, the returning controllers were aghast at the pressure and chaos of current working conditions compared to the conditions that, more than a decade earlier, were considered severe enough to motivate a general strike in which safety issues were a major area of concern. Although these more experienced controllers probably have a wealth of experience that could be tapped, by and large they cannot adapt to new working conditions. This is a case where knowledge acquisition techniques could be used to derive the particular historical perspective of the veteran controllers.

Ability to articulate actual practice

Some of the most valuable informants in a domain are those able to analytically describe *actual* practice (as opposed to official or stereotyped reports of practice). In some cases this is difficult to achieve due to lack of ability to articulate on the part of informants. In other cases it may reflect the nature of the performance patterns of the informant community. In the case of Trauma medics, it was discovered that there was a lot of value in performing field observation of medics in action to cross-check and elaborate on the basic interviews. Because of the split-second pace of decision

making in the field, it is difficult for trauma medics to fully describe what they do, as much of their behavior is learned reaction, and difficult to describe outside the practice setting and to those unfamiliar with the domain. TCIMS also benefitted from access to practitioners who were qualified instructors.

Reflectiveness /Introspectiveness

The ability to reflect upon one's expertise and analytically report the results of that reflection is affected both by the nature of the domain, the personality of the informant, and the structure of the training in the field. The experience in TCIMS indicated that trauma medics are not ordinarily introspective about their practice unless they also happened to be instructors. The hypothesis offered is that this is largely a side effect of the time-critical nature of their work as well as the training to react quickly to situations.

4.1.3.4 Selecting and Characterizing Artifacts

Many of the issues in selecting an artifact for study parallel the issues involved in characterizing informants. Issues of outdated knowledge are practically identical, though an artifact might not be as insulted to be considered out-of-date as a human informant. Since artifacts are workproducts in their own work setting, they share some features in common with the workproducts of the KA setting itself. Here we outline the features of artifacts that are particularly interesting from the point of view of knowledge acquisition.

Reflection

It might seem odd to think of an artifact as being more or less reflective than another, given that artifacts are inanimate objects, and thereby do not think. However, artifacts can display varying degrees of reflectiveness in their content. Some artifacts relate directly to a single system, and do not attempt any degree of generality about what they do as a member a class of capabilities. Examples of such artifacts are user's guides to software packages, or executable code itself.

On the other hand, some artifacts explicitly try to place the system they describe into a larger context, either by comparing it with other similar systems or describing it in more general terms. Examples of this sort of artifact are survey articles written by focus practitioners and translation packages that adapt the interface of one system to the standards of another.

It is not necessary for an artifact to refer to several systems to be reflective. Rationale documents, which document why a system was built the way it was, are examples of reflective documents that refer to a single system. Such documents typically relate a system to some broader design concepts, and project how the design will behave when it is adapted to new uses.

Accessibility

For humans, accessibility usually has to do with the demands on the human's time. In the case of artifacts, accessibility is often a matter of politics. Some artifacts might be protected by commercial agreements (such as consortium confidentiality), while others might be classified by a national government (often the case in military projects). Some artifacts will cost money (e.g., the executable code for a commercial system), which might be beyond the project's budget.

Intended audience

As is the case with knowledge acquisition workproducts, artifacts typically have an intended audience, which affects the notation for the representation that was used, the terminology, and even the accuracy of the statements. A number of commercial systems have a great deal of literature written about them. Some of these books are intended as "beginner's guides", and hence do not reflect all the complexity that might be inherent in the system. Some beginner's guides might even present factually incorrect simplifications for the purpose of protecting a naive audience from overwhelming complexity. Evaluating the context in which an artifact was intended will usually require some *context recovery* (i.e., identifying and making explicit the assumptions embedded within the software artifacts that come from the cultural context in which they are used.)

4.1.4 Assigning Representational Notations to Audiences

In Section 5.0, we catalogue a number of notations and some of the uses for which each is particularly suited. For the overall enterprise, it is useful to determine, for each audience in the project (as determined in Section 4.1.3.1), which notations are appropriate for representing knowledge for that audience. In particular, certain notations that are familiar to computer scientists (e.g., flow charts) might be unfamiliar to medical personnel. Other notations might well be familiar (e.g., hierarchical tree diagrams), but have a different meaning. A catalogue of notations to be used in the project, and the audiences for which they are intended, will make planning sessions much easier. In TCIMS, information of this sort was encoded in a template for a session report, with boxes to be checked for each notation used.

4.1.5 Initializing Dossier Infrastructure

All of the information determined at the enterprise level can be used as an index for the dossier of the workproducts to be produced by the KA project. In particular, the different audience communities, the different focus communities and the representations types are important indices for the workproducts to be created. Section 6.0 gives a detailed description of how one can build a dossier, and starter sets based on the enterprise planning components for building an index.

4.2 Planning a Thread

A thread is a lifecycle of a single entity, either informant, investigator or artifact. Thread planning is probably the most problematic planning task in Canvas. By definition, this lifecycle is not decided once and for all at its beginning in the project, but develops based on unpredictable outcomes from each session which is part of the thread.Thread planning, therefore, involves determining which aspects of an element are to be tracked, so that further planning can be performed based on the results of sessions. Here we give a list of potential risks, and information that should be tracked to help manage those risks.

Bias management

Bias is perhaps the most important aspect of a participant in the KA process to be managed. The most important biases are those of the investigator, which will cloud the information that he acquires, either from artifacts or through interviews with informants.

Bias can be controlled by taking into account what information a given KA participant has received at a given point in the participant's thread through the KA enterprise.

Example. In the TCIMS project, videotapes were made of the interview sessions. After the session, the information was written up in a report, which would go into the dossier. The reports could well be written by people who were not present at the interview itself, by viewing the video. The question of bias arises when we decide whether the person who was present at the session should brief the other team members before they watch the film. On the one hand, the briefing could clarify issues that are not on the tape, such as whether the informant was interrupted from another task when the investigator arrived, or whether he was waiting, apparently idle. On the other hand, the briefing could bias the viewer towards certain interpretations of the events, causing him to miss others. A similar problem arises when deciding whether the film should be watched in a group or separately.

This trade-off can be generalized to any KA workproduct setting; in principle, the investigators can examine workproducts in parallel or in sequence, and can consult one another's commentaries or not. Consulting earlier interpretations can save time and allow for deeper study of detail, while parallel uniformed viewing results in a broader range of interpretations.

Repetition (asking informant same question twice)

The investigators can make nuisances of themselves if they continually ask the same informant for the same information, or ask for inappropriate information. The means of managing this aspect of the informant thread is to keep careful track of what information has been gathered from which informant, and to consult this information before planning a session with that informant. Repetition management might conflict with bias management, since it might be deemed appropriate for the investigator to remain ignorant of some result of the previous interview, for purposes of controlling bias.

Interview consolidation

Closely related to the management of repetition is the management of the overall interview process. If several investigators are interested in related topics, all of which are available from a single informant, it is in the interests of the project to consolidate these interviews. This requires similar infrastructure as management of repetition, only for the information requirements of the investigators as well as the results.

Investigator as ambassador

It is often the case that the investigator comes to know so much about the focus community, that members of the target community will bring questions to him, rather than go to members of the focus community. It is possible to plan to develop one or more of the investigators to play this role; in this case, the investigator should either begin with a familiarity of the focus community, or should be involved in KA sessions in which members of that focus community are treated as knowledge sources. This strategy has the advantage that the answer person has access to knowledge that was never codified. It has the disadvantage that it discourages the comprehensive codification of the knowledge, thereby weakening the value of the dossier.

KE as educative for investigator: designing training into the process

The investigators will begin naive and will become more and more informed about the focus community as the project moves on. This means that the criteria listed in Section 4.1.3.2 will change as the investigator progresses. This can be used to the advantage of the project, as a means of bringing in new investigators after the project has begun. The new investigators can be used for sessions in which it is deemed advantageous to have an investigator who is unfamiliar with the

domain, while the more experienced investigators are used for those sessions where familiarity is called for.

KE as knowledge creating: fostering informant reflectiveness

As an informant moves through the KA process, she will be encouraged to think about what she does in her work practice, and will be exposed to the possibilities afforded by a knowledge acquisition effort for codifying, analyzing, summarizing and comparing information. Informants who are so inclined may become more reflective as this process continues, thereby changing their categorization according to Section 4.1.3.3.

Contingencies and emergent discoveries

The best laid plans will be foiled by real-world contingencies; in thread planning, which is by nature a speculative process, this is even more true. As the project progresses, surprising and serendipitous discoveries might be brought about, by virtue of the new combinations of expertise that are brought together by the KA project. Individuals may turn out to have proclivities that were not anticipated (e.g., an informant turns out to have a "pet theory" about his domain that has never been published, but when it is turned up in the KA context, turns out to win surprising buy-in among the other informants). Such events cannot be planned for, but they can be anticipated, by keeping careful track of the assumptions guiding each thread's development.

Resource management

As the threads interact in a large project, there will be problems of resource management. Some investigators will have developed the background to examine certain artifacts or understand certain informants, while others will not. The management of a thread will have to take these resource allocation problems into account, and perhaps develop a thread in a particular way so as to minimize bottlenecks later on in the process.

4.3 Planning a Session

Within the overall context of a KA enterprise, a number of individual KA sessions will be conducted. Although some global decisions will be made as part of broader KA enterprise-level planning, and some guidance has been done for the particular threads, the most detailed decisions need to be made on a session-by-session basis. This section will mostly discuss issues concerned with sessions involving interactions with informants, as the planning issues surrounding artifact analysis are somewhat simpler.

An overall "architecture" to a session can be defined, which involves up-front planning and scheduling of the session, preparatory work for various participants, the primary elicitation event itself, the "write-up" or codification process, validation of the documented results as required (though this could spill over into another session context) and dissemination of the results, via the dossier, to other participants in the KA enterprise. The eventual use of the KA workproducts by people in the target community (e.g., technologists viewing interview reports) is not here considered a core part of the knowledge acquisition process proper. Planning aspects of these various elements of the KA session are discussed in the paragraphs below.

4.3.1 Preparation

Planning a KA session first and foremost implies making decisions about the basic elements of a session, including the following:

- The session objectives and topic(s) of focus: What is the primary purpose of the session? What topics are of interest? Who will be the primary audience for the resulting codification of session results?
- The setting(s) of interest: Is a particular setting going to be the focus of the session? Might there be parallel sessions planned to elicit similar descriptive data about different settings? Or is the setting being used to provide a "normative" view of domain-specific practice?
- The investigator(s) conducting the session: Who is available, and best qualified, to conduct a particular session? How will that session develop the threads for the various investigators involved? Is the intention to provide them focused experience in a few topics or settings, or to expose them to a diversity of perspectives?
- Informants that might be involved in the session: Who is available, and who best suited to provide information on the topics of interest?
- Inputs: What artifacts or previously developed KA workproducts might provide prior background information for the session? Which could be used directly as knowledge sources and/or a basis for interactions during the session itself?
- Results: What are the anticipated KA workproducts to be produced from the session? What formats or representations should be employed?

Informant preparation

Where is the informant in his thread? Has the informant received project orientation/familiarization? Has he participated in previous KA sessions? What additional information does the informant need to know before hand? Will he be asked detailed questions about an area in which he may have spotty recollection? Will he want to do some review work before the session? Or is this something specifically to be avoided? One KA objective (or guidelines) may be to ascertain "knowledge in use" rather than the official knowledge formally available.

> Example. In interviewing a helicopter transport pilot for TCIMS, the interviewers asked about an equipment maintenance checklist that hangs on the wall in the office and is used to monitor scheduled maintenance for the helicopter. The informant mentioned some example categories of maintenance tasks from memory in describing the sheet. Different data would have been obtained if this topic of interest had been previously called out in the interview preparatory material. The informant might have reviewed the material to give a more comprehensive description, might have been more cautious if there are legal or contractual restrictions on "correct practice" that are sometimes relaxed in practice, or might have simply told the investigators to go look at the document itself.
>
> In this situation, the un-prepped responses yield possible data about "foreground" versus "background" maintenance problems, or "typical" versus "atypical" problems. Naturally, there is a danger of over-interpreting. The choices made by the informant about which maintenance problems to mention might be related to relative frequency of occurrence, relative criticality of the problem, the length required for a fix with its concomitant impact on equipment availability, the informant's expectation of what examples would be understandable to a non-mechanic, or any number of other factors.
>
> Thus validation of some sort would be necessary. This could be done within the same interview, in a follow-up interview, in a request for comment included with a transcript submitted to the informant for review, or by cross-checking in interviews with other informants. The interview results could also be cross-checked against the artifact itself. In this case it would be of particular interest to view the artifact in context to see how it is positioned, what mar-

ginal notations are made, etc. Theoretically results could also be checked with a maintenance mechanic. However, here the question of domain focus becomes relevant. Since the overall objectives of the TCIMS project are focused on emergency trauma care, the transport domain is itself only tangentially related; the domain of helicopter equipment maintenance would be an even more remote "second cousin" and probably out of scope for the KA effort.

It might be desirable to let the informant know what kinds of questions will be asked beforehand, or to allow him to prepare some answers to general questions, to get the session started. Or this might be intentionally avoided, so as to avoid having the informant look up official answers, rather than giving his own.

Example. In the TCIMS interview of the helicopter pilot, a template was given to the pilot before the interview, describing the nature of the material to be collection (task information), and asking some general questions about major responsibilities, durations, etc. Any answers given on this form were used as starting points in the discussion. In the case where no answers were given, the interview would begin with similar questions from the investigator.

Investigator preparation

Where is the investigator in her thread? What information pertaining to the session topic could be acquired from basic sources, including well-known texts or other KA session reports? What information could bias the investigator inappropriately for the session?

Example. In the TCIMS interview of a helicopter pilot, the investigator brought with her knowledge of the structure of the emergency room and hospital for which the pilot worked; in particular, the administrative structure of the transport service (as a private contractor, hired as a company by the hospital) had considerable impact on understanding the impact of rules and regulations governing the work. This information had been gained earlier in the investigator's thread by her participation in other interviews with personnel from the emergency room transport service.

In contrast, it would not have been advantageous for the investigator to have begun the interview with detailed knowledge of helicopter flight, since the topic of interest in the interview was the responsibilities of a medical transport pilot. Knowledge of helicopters could have biased the discussion toward details of the mechanics of flying, rather than the interactions with the medical constraints of the situation.

4.3.2 Session Performance

Full exploration of the many subtle elements involved in conducting interviews, joint design sessions or artifact analysis are beyond the scope of this guidebook, which is oriented primarily towards the planning and dossier management aspects of KA. However, some aspects of performance of the KA session very directly reflect the key Canvas principles.

Rapport

Building rapport is an essential part of any productive human interaction, but in KA it is critical for developing the trust needed to validate information and enable mutual knowledge building and discovery. If relationship is the end and trust is the vehicle, then rapport is the starting point. Planning and conducting sessions so that rapport can happen and is given appropriate emphasis at different points during the session is an often-missed and undervalued aspect of KA planning.

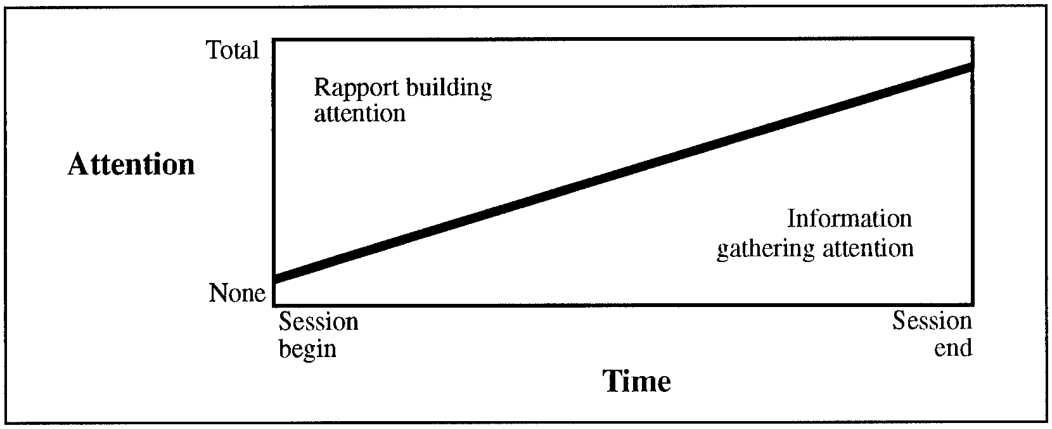

Exhibit 7. Attention in a Session Devoted to Rapport Building versus Information Gathering

The beginning of the session is most critical, perhaps the first 20 minutes. In these first minutes, the researcher will benefit from damping his curiosity about facts and content, while focusing on the person's background, the physical environment, and nonverbal communication cues. For example, this usually means emphasizing eye contact over note-taking. In short, it means making a connection first and foremost.

One way of thinking of a KA session is illustrated in Exhibit 7. The picture shows that early in the session attention to exchange of information is low; development of rapport is high. They become equal roughly half way through, and shift toward the end. When sessions are going well, they equalize about one-third of the way through, shift toward information exchange about two-thirds of the way through, and the final third opens up the possibility of joint discovery, knowledge creation, and possible knowledge modeling.

Thinking in terms of these proportions can be significant session design considerations. For example, some researchers may need to plan a set of more personal questions for the opening. Key or controversial questions should be saved until rapport is clearly being established. Pictures or visual models should not be introduced too soon, breaking eye contact and connection between informant and researcher. Finally, ending with plenty of time for creative exchange to happen spontaneously should be planned. Sometimes it is just as the session is ending that an informant feels greatest trust and wants to give most. These are just examples of how session planning can be enhanced by considering the structure of rapport building.

Session logistics

A number of logistical problems can arise in planning a session, including gaining access to the informant at all (in military settings, it is important to observe the chain of command to assure that access to the informant will not be denied at the last minute), planning a location for the meeting, arranging for any equipment (recording devices, demonstration machines), scheduling, etc. It is worthwhile to consider why logistics decisions were made as they were, so that if something goes wrong (e.g., a technical problem with a video recorder), a timely decision can be made about how to react.

Context setting

The informant must be oriented as to the nature of the KA enterprise and the larger project of which it is a part. This could be done as part of the preparatory activity for the session, but it is a

good idea not to make assumptions that any preparatory material has actually been reviewed thoroughly. For example, for one interview which we observed, the material sent in advance for the informant's benefit had never been passed on. This might be typical of expert informants in high-pressure, time-constrained environments.

Open versus closed session facilitation

There are trade-offs involved in the session planning process. For example, insisting on a strict focus for an interview may help keep the interview on track, or may blind the investigators to essential information volunteered by informants which happens to be out of scope of the session objectives. The session objectives should help clarify this. If the purpose of the session is baseline familiarization for the investigator, or getting an informant "on board" as an interested contributor to the project, more flexibility might be appropriate in selecting an open versus closed style for the interaction.

Writing it up

Perhaps the most difficult part of a knowledge acquisition session is the write-up; this is the work-product that will be entered in the dossier and serve for some audience community to know what happened during the session. One of the major problems is to ensure that as much of the information that was elicited during the session as possible makes it into the dossier; this can be simplified if the session itself followed strict goals. However, as mentioned above under Open versus closed sessions, there are other reasons for the session to be less structured.

In writing up the session, it is important to keep in mind the audience for the write-up. If the write-up includes any representations, these must be expressed in a notation that is accessible to this audience. Notations were identified with their target audiences in Section 4.1.4; this information can be used as a guide in selecting notations to use in write-ups. There might be need for several write-ups, for different audiences, in particular, one for the investigator team, to help them gain understanding of the dossier, one for the informant, to verify the information, and one for the target audience.

Validating a session

Interviews with informants should be validated in a special session with that informant. Validation of artifact studies might involve consulting an informant, though sometimes a review by the investigating team might be sufficient. This could be as simple as having the informant review the write-up, or as complex as having a follow-up session to examine the details of the critique. The trade-offs here involve the informant's time and commitment to the project.

A particular difficulty arises in validating representations of variability, not only because of unfamiliar representation notations, but because such models do not represent the "world view" of any one informant who could validate them. Special validation techniques may be required. For example, multiple informants from different settings could be brought together to validate a model that captured the range of variability in the knowledge gathered from those settings. Applying such techniques may have the side-effect of changing the way informants organize and reflect on their own domain knowledge. In this case, the KA and modeling process is a definite intervention in the dynamics of the work setting, whether or not the final result is introduction of new technology into that setting.

4.3.3 Variants of Sessions

So far, we have treated sessions as if they were controlled interactions between a single investigator and a single knowledge source. In fact, a single session could manage an interaction among several actors in the KA project. Many of these types of interactions are familiar, either from everyday work practice, or well-known knowledge and requirements engineering practice. The Canvas framework applies just as well to these types of sessions. In fact, a healthy mix of different session types is likely to result in a more flexible KA effort.

Walkthroughs

An investigator studying an artifact alone will often have several questions about the context in which it is used, the meaning of the terms, how it relates to other artifacts, etc. One way to handle such a situation effectively is to plan a session in which a human informant works with the investigator as he studies the artifact, and "walks him through" the various pieces. This is particularly useful when the artifact has a process nature, such as a procedures guideline manual.

Demonstrations

Closely related to the walkthrough, especially given, as described in Section 3.3.2, that knowledge acquisition will often occur in technology intensive settings, is the program demonstration. In this case, the artifact to be studied is a piece of software, which is run during the demo, and is accompanied by some explanations from an informant. This format provides a great deal of possibilities: multiple investigators (as in a demo performed for a group), specific questions and answers, and limited experimentation on the part of the investigator. Experimentation can be either planned or not. For example, during a demo of a system that diagnoses engine problems in an automobile, the person giving the demo might ask the audience, "Think of a problem with your car." Or, if the investigator is in a position to drive the demo, she might ask, "What would happen, if the program were to receive this input?". Demonstrations can place practitioners into a context with which they are familiar, thus facilitating knowledge acquisition.

Automated sessions

An extreme example of a knowledge acquisition session is the fully automated session, in which a human informant is left alone with a computer program that leads him through the session. Manusco and Shaw [21] report that an automated session is often liberating for an informant, because the informant has a better feel of ownership over the resulting knowledge. She also reports an increased sense of trust, that the machine will not judge the completeness or coherence of the knowledge. Weizenbaum [53] reports the uncanny success enjoyed by the Eliza program in gaining the trust of its users, even to the point that they would ask other people to leave the room while using it. Although this is a promising approach, the current state of the art in fully automated knowledge acquisition is such that only a limited set of highly specialized representations can be produced in this manner.

Other automated sessions include automatic processing of documents, such as word frequency counts, which provide objective and sometimes surprisingly insightful information about natural language documents.

Wizard of Oz sessions

A common method used in human computer interaction studies, expert system design, and requirements engineering is the so-called "Wizard of Oz" method, in which the investigator pretends to be a computer system that will be introduced into a workplace. The informant deals with

the expert as he would with the system. The behavior of the system can be changed simply through agreement between the investigator and the informant. Although the method has been traditionally used to project the impact of new technology on the workplace, it can also be used to uncover the hidden assumptions in the work practice.

Observation sessions

A very useful, though sometimes demanding method for acquiring knowledge of a work setting is through observation, that is, the investigator is present at the site of the work practice and observes it in action. One possible problems stem from the fact that practice is likely to change under outside observation. The solutions to this problem include having the investigator enter the work setting in some accepted role, often as an apprentice or junior colleague, and make observations while actually performing in the work practice in the focus community. Another solution is to provide automatic surveillance, through video or audio recording machinery. In technology dependent settings such as those described in 3.3.2, the technology itself can provide opportunities for automatic observation.

4.3.4 Issues in Session Planning

A plan for a knowledge acquisition session has to deal with a large number of competing constraints and trade-offs. These issues can affect any aspect of the session plan, including the structure of the session, the participants, and the representation notation used for the write-up.

Dynamics of single versus multiple informant interactions

Consider the decision to interview a domain expert in a one-on-one setting versus within a group. Canvas provides guidelines for identifying some of the key issues that should be considered. Social interactions will occur between multiple informants. This could result in additional information emerging from the interaction which could be of value to the KA enterprise. If the multiple informants are co-workers from the same setting, anecdotes and recollections of past experience may come up as part of the interaction that an external investigator would have lacked the context to introduce. These particulars might yield more general principles relevant to the topic of focus. This reflects the principle that important types of knowledge are carried in the social interactions that make up the dynamic aspects of any community of practice.

However, communities have defined social relations, and these relations will affect what kinds of knowledge emerge from the interactions and what aspects of this knowledge people will be willing to share. If there are line-of-authority relations among the informant group those in a lower staff position might be unwilling to voice opinions that would be deemed critical of the management, whereas in a one-on-one interview they may speak more freely, especially if guaranteed confidentiality or anonymity of attribution. Conversely, those in positions of more authority might be loathe to express uncertainty about future business in an exposed setting.

From the thread perspective, an investigator's level of experience and degree of credibility within the focus community (established over a series of interactions) would need to be considered. A multi-informant knowledge elicitation session runs the risk of turning into an interaction between the informants, and the investigator might need to be prepared to step into the role of mediator or facilitator to some extent. From an informant's thread standpoint, it might generally be a good idea to interview an individual first, in order to predict how his or her interaction styles might affect others in a KA setting. On the other hand, a group interview with very clearly stated goals and structure could be used as a screening device to earmark promising candidate informants for further work.

Evaluation of investigator results

Work products created by investigators may be subjected to multiple types of review. One of the most valuable is review by the informants themselves, with attention to accuracy and completeness. Further review may be carried out by peer investigators, with consideration to clarity, attention to established investigator processes, suggestions for follow-on KA activities. Another important form of review is by an established group of experts as in TCIMS that takes a more global look at the body of KA results and the credibility of the sources of the information. Finally, there may be reviews by technologists with respect to the usability and formality of the information, its suitability as source material for work products such as software requirements and design documentation.

KA enterprise input to session planning

Many of the decisions made while planning the enterprise can be used as guidelines in planning the individual session. We will demonstrate this with an example from TCIMS, where stakeholder issues, the planned audience groups and their associated notations for representation (see Section 4.1.4) all have an impact on the successful planning of the sessions.

> Example. Two lessons learned from the TCIMS project concerned the accessibility of information to various target groups. On the one hand, the group of technologists (i.e., the system developers who will be introducing new technologies into the field) were reluctant to recognize either the difficulty of the knowledge acquisition effort, or, more importantly, its value to their task. Technologists complained that there was too much verbiage in the material, and that it did not tell them what they needed to do to develop their systems. On the other hand, notations used in some KA products were unfamiliar and perhaps even alien to the medical community. The effect of this was to increase the difficulty of model validation.
>
> Both of these types of problems can be ameliorated, or at least better anticipated, by paying close attention to the intended audience of each workproduct. This would both reduce the amount of information that the technologists would have to deal with, as well as make sure that the information that they do access is understandable to them. Similarly this would remove the necessity for members of the medical community to learn notations that are unfamiliar to them.

The lesson to be learned for the knowledge acquisition project planner is that it is necessary to explicitly know the intended audience of each KA workproduct, and to tailor each workproduct for its intended audience. In particular, a KA workproduct that will be accessed by multiple audiences may need to be maintained in multiple forms. Here, automated tools to support seamless transfer of information among representations would clearly be a boon, as the collaborative nature of the KA enterprise means that changes can ripple back and forth along the life cycle from raw data acquisition to more formal modeling and interpretation of the data.

Variability issues

In general, in dealing with a given informant, we can distinguish between variability that can be directly elicited from his experience and variability that can be modeled only by aggregating or integrating his experience with other data. More specifically, we can distinguish the following cases:

- In some cases, particularly reflective or expert informants can offer data that is already close in form to a domain model of commonality and variability for their domain: an abstraction of their experience. Or, in many stable domains, elaborate codifications of data are already available (e.g., in the medical domain, established protocol or procedures documentation.) In

general, such data needs to be handled in the KA process as an artifact, but as a particular kind of artifact which provides secondary (second-order, interpretive or meta) data as it occurs within the focus community.

- Variability that is part of regular work practice for the informant but is articulated through or as a result of the KA interaction. This could be documented as part of a work process model, procedures (with contingency conditions,) etc. These could range from truly routine, low-skill activities to activities requiring expert judgment. For a practitioner such as a medical staff person, triage procedures are an example. Here, the knowledge is codified as a result of the KA process but was known or accessible to the informant prior to KA participation.

- Variability can be elicited from informant's experience through a process of reflection. In this case, the informant has the varied experience (moving jobs, working at different sites, etc.) to offer the observation of the variability. However, the KA interaction has triggered a reflection or learning process that changes the informant's relationship with their own knowledge or practice.

- Variability observations result from making new information available to the informant and thus sparking reflection that is a direct intervention of the KA process. This could include walking an informant through a set of artifacts similar (but not identical) to those with which he is familiar; bringing together a group of informants filling similar job roles to compare notes; or asking an informant to validate a KA workproduct which represents codified variant data gleaned from other informants. As with the case above, the learning involved in sparked by the KA interaction but in this case more of an intervention has taken place, since new data is made available to the informant.

- Variability is observed by the investigator by synthesizing data obtained from multiple informants, artifacts and/or settings. This case is often the assumed scenario in KA; that is, representations of variability are assumed to be products of the investigators and not directly of the informants.

  Example. Two medics might have different ways of classifying the various triage decisions that they face. In this case, we must compare not only a set of triage instances or cases which we then synthesize into our own classification or comparative model, but also the way the two medics have differently "made meaning" out of the range of cases in their respective bases of experience.

The examples above range over several different factors. The primary factor is the extent to which the knowledge is already extant in the focus community, is created via interpretation by the investigators, or emerges through collaboration between informant and investigator. The KA plan should make necessary distinctions between these different cases and track representations differently depending on their origin. Otherwise it is difficult to correctly interpret the representations or models. Each such model, whether a previously created informant model treated as artifact, or a KA workproduct created in collaboration with investigators, embeds a "theory" about variability in the domain. There may be multiple such models, or a single such model may diverge from the theory that emerges from the investigators' own models in the same area. This becomes a model resolution issue that reflects variability at a meta-level from that observed in the work practice of the domain itself. Some social scientific academic work has been done in relevant areas such as meta-ethnography that deal with the synthesis of multiple, qualitative interpretive data analyses [25]. In general, though, this is an area requiring substantial further research.

Notations for representing knowledge acquisition need mechanisms for describing variability. These will be discussed in more detail in Section 5.0, which discusses representation strategies for KA.

Investigator bias

The investigator's previous experience and degree of relevant expertise can be a significant factor to consider in planning. Bias on the part of the investigator can affect the accuracy, consistency and completeness of elicited knowledge. The effects of bias can be anticipated during enterprise planning, and the development of the bias itself can be tracked during thread planning. The effects listed below can be counteracted to some extent during session planning:

- Missing data (did not ask the question).
- Inaccurate data (misheard, misreported, imposed own view).
- Interaction led to false data (led the witness, created resistance, or created an over-willingness to please and give the "right" answer).
- Mixed source of data (uses session as vehicle to interject own opinions and knowledge without acknowledging this).

Strategies to counteract risks of investigator bias include the following:

- Having informants interviewed by teams that include an investigator with a high degree of topic expertise and a relative novice (with respect to the topic).
- In Section 4.1.3.2, the investigators were characterized according to their skill levels in specific techniques and procedures for suspending unwanted influence of their expertise on a session. Selecting investigators with these skills for sessions where the effects of bias are particularly risky or likely.
- Build into the session structure explicit solicitation of feedback from informants about their perception of how well the investigator has performed their task. Were their remarks properly understood and noted? Were important points missed or glossed over?

Informant bias

Bias on the part of the investigator can skew results in various ways, as described above. Bias on the part of the informant is slightly different in nature. Of course, an informant will be presenting personal opinions and viewpoints; this is a given in dealing with the knowledge source. The problems here can take the following forms:

- Lack of clarity in distinguishing personal knowledge and opinion from consensus knowledge of the community of practice.
- The "say versus do" problem: what is reported may not correspond to actual practice.
- Unwillingness to share knowledge, or distortions ranging from "white lies" and politically correct answers to deliberate falsification.
- The informant may have a lot invested in the uniqueness of their knowledge and problems; or conversely may be too ready to report in generalities and "normalize" practice.

Different strategies for planning sessions can adjust for some of these factors. Relevant factors in selecting the right strategy include the closeness of the session's context to the work setting about which knowledge is required, and the degree of awareness of the informant that knowledge acquisition is occurring. These various strategies can be viewed as a kind of spectrum of options:

- The most typical interaction between an investigator and an informant would be in a formal interview outside the work setting (e.g., in a conference room, perhaps with a tape recorder or video, certainly with the investigator taking notes during the session). If the informant is being asked to recall detailed procedural knowledge this may be an awkward setting. (The scenario elicitation process is one attempt to facilitate recall by appealing to natural ways of conceptualizing work flow and task boundaries.)

- An interview conducted in or near the work setting itself may improve the closeness of fit between the reported and actual practice. Passive observation of work actually being performed, with the investigator visible and acknowledged but otherwise not directly interacting, would generally yield richer data along these lines. However, the activities in the work setting will be affected by the presence of an observer.

- Participant observation, where the investigator plays a work-related role within the setting of focus, can be less obtrusive because there is an acceptable role for the investigator within the normal roster of work-related roles. This personal, experiential data can result in far richer knowledge transfer. How much of this is effectively *codified* depends on the skill of the investigator.

- Non-intrusive data collection (e.g., instrumented tools for data collection and logging) would in theory enable obtaining more accurate information about real versus official practice. Here an ethical issue arises. In theory, clandestine observation would yield the most objective data, yet it would generally be unethical to collect and utilize data on this basis as part of a knowledge acquisition effort. When informants are aware and willing to participate, non-intrusive data collection has the advantage of not "breaking the flow" of the event.

5.0 Representation of Knowledge

Representation of knowledge forms the cornerstone to knowledge acquisition. Any knowledge that will be acquired, whether it is formed collaboratively with the informant, or learned by rote, whether it is intended for the focus community, the investigator community, or some other community, has to be represented to be passed on to its audience. The details of the notation used to represent the knowledge constrain what knowledge can be acquired; if some knowledge cannot be represented in the chosen notation(s), then it cannot be acquired.

Section 5.1 presents some fundamental attributes that can be used to characterize notations themselves, and will apply to any representations developed in the notations. In Section 5.2 we describe four types of notations that can be used for knowledge representation and describe their characteristics in terms of the attributes presented in Section 5.1. Section 5.3 will present conclusions about choosing notations to represent knowledge.

5.1 Attributes of Notations

There are fundamental attributes that can be used to describe the *expressive power* of a notation. This expressive power of a notation determines what knowledge can and cannot be represented using the notation. The attributes can be used as criteria to perform trade-off analysis to choose a notation that is suitable for capturing a particular type of knowledge.

5.1.1 Dynamic versus Static Information

Notations vary in their abilities to represent dynamic and static information. Processes and algorithms are dynamic in nature, and require very specific notations to capture this dynamic aspect. Category structures and lexicons, on the other hand, do not have a dynamic aspect to them, and have another set of notations altogether. Dynamic and static in this sense do not refer to how quickly the knowledge changes, rather, to the nature of the thing to be explained itself.

Flow charts, instruction booklets and programming languages are common examples of dynamic notations; data dictionaries, grocery lists, and class structures are common examples of static notations.

5.1.2 Variability versus Commonality

How we deal with variability in a knowledge acquisition project is of fundamental importance in Canvas. A related issue is that variability can only be dealt with to the extent that our notations can express it. When a notation is capable of expressing variability, it means that it can express information about a set of phenomena, explicitly indicating what the exemplar phenomena have in common, and what they do not have in common. Population statistics (e.g., 35% of supporters for the candidate live in cities, 60% in rural areas) are a common example of a notation that is used to represent variability (e.g., cities versus rural areas). A particular notation can be good at representing either commonality, variability or both; these variants, including the perhaps counter-intuitive idea that a notation might be good at representing variability but not commonality, are treated in detail in [38].

5.1.3 Knowledge Types

There are a number of ways to classify types of knowledge. Here, we are interested in types of knowledge that impact how easily the knowledge can be elicited from an informant, and what impact the types have on the representation of the knowledge. The list below is intended to give a rough idea of types of knowledge as they relate to knowledge acquisition. Notations used for knowledge representation can be rated as to which knowledge types they can easily be used to represent.

Procedural knowledge is a type of knowledge which consists of the skills or tasks a person can accomplish or perform. This is what an informant would describe as "knowing how" to do something. Procedural knowledge is often an automatic response to stimuli, can be reactive in nature, and is therefore difficult for an informant to explain or describe. An example of procedural knowledge and the difficulties inherent in acquiring it can be shown by considering how difficult it is to explain how to tie a shoelace.

Declarative knowledge is a type of knowledge often referred to as "knowing that". It is surface level information that an informant knows he possess and that he can easily verbalize. An example of the difference between procedural and declarative knowledge can be seen in a shoe tying example. In contrast to the procedural knowledge of how to tie a shoelace, the declarative knowledge that a shoe has two laces, that there are two types of knots that can be tied (i.e., granny and square), is easy to access.

Semantic knowledge represents one of the two types of long term memory. Semantic knowledge includes words, symbols, facts, and definitions. Tapping into this knowledge source is critical for effective knowledge acquisition. When the goal of the project is to support software system development, semantic knowledge can be used when a system must accurately emulate a task formerly performed by a human. Again using the shoe example, the definition of various shoe types (pumps, heels, Birkenstocks, etc.) is an example of semantic knowledge.

Episodic knowledge is the second form of long term memory and includes autobiographical and experiential knowledge which an informant has grouped or "chunked" together. It contains information about temporally dated events and the temporal/spatial relationships among these events. An example of episodic knowledge would be driving a car to work. Although one could explain the route taken, it would be difficult to explain the strategy of changing lanes, knowing how the car operates, etc. This type of knowledge is deeply ingrained and difficult to access. This explains why we can arrive at work and not remember any part of the journey.

5.2 Example Notations

The choice of a favorite notation to use for representation is probably the most personal choice made by an investigator in a knowledge acquisition project. Someone who participates in many such projects will probably develop a few notations with which she feels comfortable and is able to use easily. We cannot attempt a comprehensive catalog of knowledge representation notations in this guidebook. We have chosen to present a set of notations that covers a wide range of capabilities with respect to the attributes given above. In any project, the set of notations that will be used will depend heavily on project context and the skills and backgrounds of the investigator team. The set of notations presented here should be looked at as an example of some of the possibilities.

The description of each notation type is accompanied by a characterization in terms of the attributes of notations given above. These characterizations can be used by the reader as examples to guide in characterizing and selecting notations suitable for a particular knowledge acquisition

Title: Parachutist's Chute did not open

Environment: Training Jump: No enemy present

Medics and buddy approached casualty after seeing him fall

Performed initial Assessment. Findings included:
- nasal fracture (congested airway)
- patient unconscious
- Bilateral fracture — legs distorted
- Pelvis probably broken

Requested Ambulance by radio
Intubation done to improve airway
Provided patient with oxygen
Prepared patient for evacuation
Performed secondary Assessment
Straightened and splinted legs
Started IV
- patient regained consciousness

Loaded patient onto Ambulance (20 minute transport)
- patient arrived at field hospital

Field hospital performed surgery

Exhibit 8. Example Scenario from SEPWeb

project. One of the main contributions of Canvas is to provide a framework where this trade-off can be managed, by combining representations at both extremes, as well as along the spectrum in between, in a single coherent dossier.

5.2.1 Scenario Notations

Scenarios are used to represent single threads of execution. Scenarios contain environmental context information, initial state information, actors, items in the environment, time referenced transitions, and a final state.

Scenarios can be used to describe specific situations that occurred and the events that took place in these situations. More generic scenarios can be used to describe the tasks carried out by a particular role. There are two types of scenarios: "As Is" and "To Be." The As Is scenario documents current practice and the To Be scenario can be thought of as a proposal for the future.

Scenarios are generally presented as text files in combination with maps or other graphic information. In some cases timelines are also included. Scenarios are decomposed into events (or tasks). Each event is listed, described, and possibly further decomposed. At the atomic level, an event is a responsibility assigned to a single entity. This entity can be a person, computer, or device. For example, "Determine Pulse" is a event that could be done by a medic or a medical device. Exhibit 8 shows an example scenario taken from the TCIMS project repository SEPWeb. Notice how it describes the particular roles that the performers (buddy and medic) and devices (radios and ambulance) play in the event.

Tell the domain expert:

We will go through the following steps

- select a scenario
- describe the scenario to me as an overview
- describe the actions that a situation assessment "person" would take at each point in the scenario
- review the tape of the above to explicate the reasoning and justifications which you used at each point.

From this I will attempt to pull out the important criteria and types of reasoning employed.

Exhibit 9. Steps for Scenario Generation

From the viewpoint of the target community, the primary uses of scenarios are to feed task decomposition (see section 5.2.2) and for system test and evaluation. From the viewpoint of the focus community, scenarios help ensure the development is focused in the right direction.

The development of sample scenarios that describe responses to specific situations by an informant allows us to look into a specific set of situations to get a clear idea of the tasks and responsibilities involved within the informant's area of expertise. By decomposing scenarios, we are able to clearly identify the key participants, identify the sequence of events and identify the key issues and communication requirements generated during the scenario. The informant and investigator should work together to determine the major occurrences of each scenario.

Providing a skeletal structure for each scenario developed ensures that the key elements are identified. Exhibit 9 shows an example of structured steps that can be used for scenario generation. Working from this set of steps, the investigator and informant further define the scenario and its attributes.

Scenarios are a way of capturing procedural knowledge in a specific situation. Like processes, scenarios describe dynamic aspects of the world. Variability in the scenario is not ordinarily an issue, as a scenario represents a single thread of execution. Variability in the domain of the focus community is addressed by composing a set of scenarios that encompass the important aspects of the domain with sufficient variations to be adequately descriptive.

5.2.2 Task Notations

Task notations allow decomposition of a task into its subtasks. Task decomposition can used to represent a task performed by the focus community that occurs repeatedly over a wide range of scenarios. For example, the task of "Casualty Assessment" occurs repeatedly in all battlefield care scenarios, although when it is performed (and by whom) differs.

Task notations are particularly useful in the selection of areas to be considered for automation (using the As Is decomposition) and in determining how automation will be accomplished during system development (using the To Be decomposition); therefore they are often the notations of

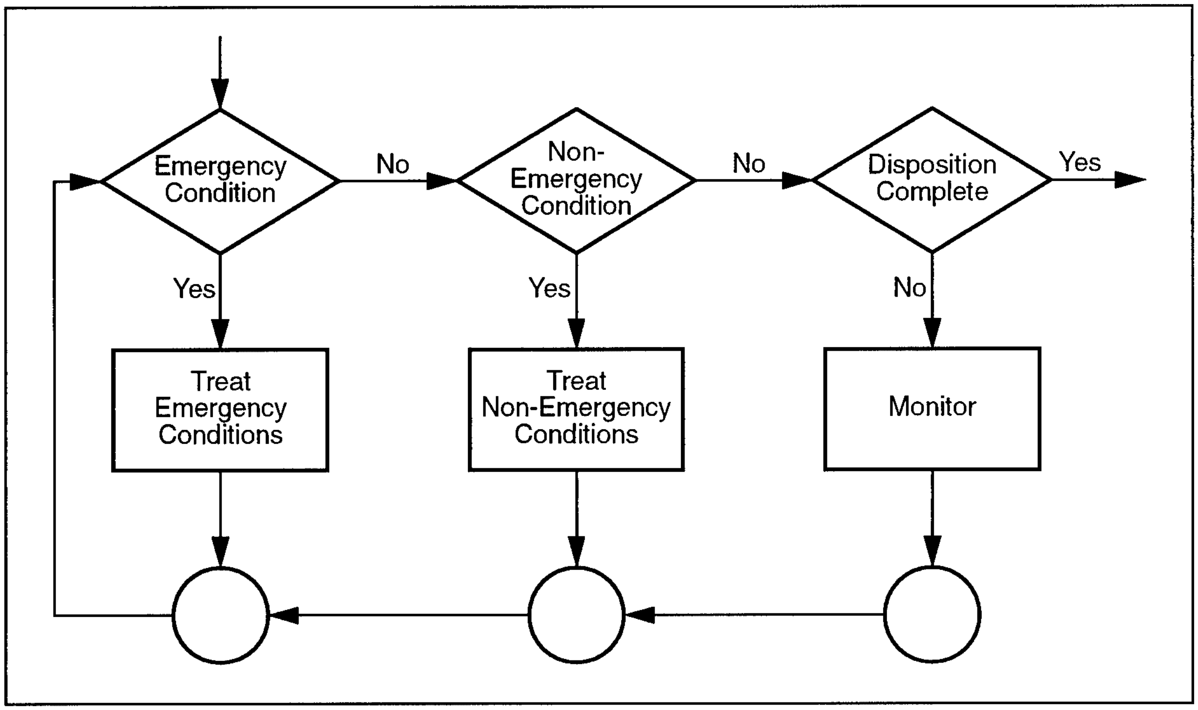

Exhibit 10. Flowchart

choice in knowledge acquisition projects aimed at technology development. Task analysis is used to identify major tasks or job functions and their relation to the overall job being performed. This type of analysis requires the investigator to be able to break tasks and subtasks down into manageable chunks that will be represented.

Task notations vary widely. They can be selected on the basis of the character of the tasks being described. A textual notation can be used to describe subtasks, timing, constraints, decisions, and resources. Graphical notations include flowcharts, modified petri nets, event trace diagrams, and state transition diagrams. Flowcharts are used for decision oriented task sequences (see Exhibit 10). When interaction and coordination is a dominant element, interaction diagrams such modified petri nets and event trace diagrams can be used. The modified petri net (shown in Exhibit 11) is a petri net with the additional features that tokens may be treated as objects that retain their identity through transitions. The modified petri net is well suited to provide a view of overall interaction and coordination, but lacks temporal information. Event trace diagrams (shown in Exhibit 12) can easily represent time, but the complexity of the diagram grows quickly as its scope is increased. As appropriate, state transition diagrams can be used to record transitions associated with performing tasks.

Task notations are versatile in the types of knowledge they can represent. While they primarily represent procedural rather than declarative knowledge, they can express both semantic and episodic knowledge. Like scenarios, tasks represent dynamic processes. The degree of representation of variability in task notations varies greatly. The event trace diagram is essentially synonymous with scenarios in that it shows a fixed sequence of events, so there is essentially no variability present except any variability in times associated with each task component. A flow chart represents the limited variability supported by decisions within the process being modeled. The inherent non-determinism of the petri net implies that any single diagram represents a broad range of behavior subject to the constraints imposed by the coordination activities, such as "Vehicle Management" in Exhibit 11.

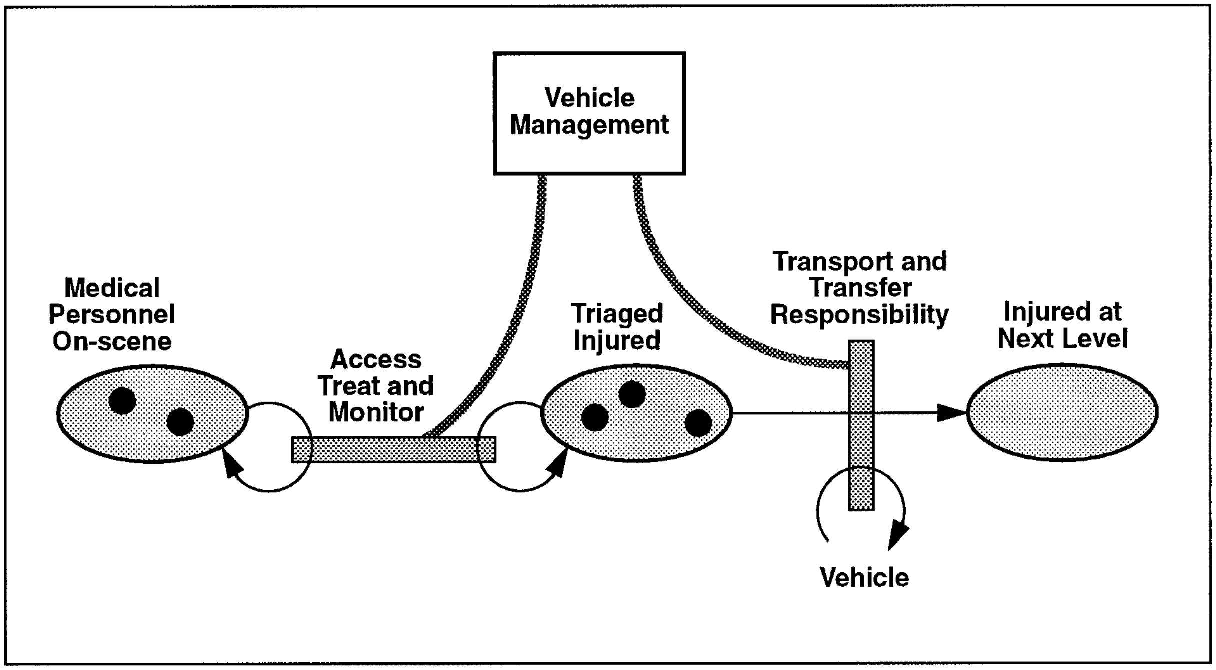

Exhibit 11. Modified Petri Net

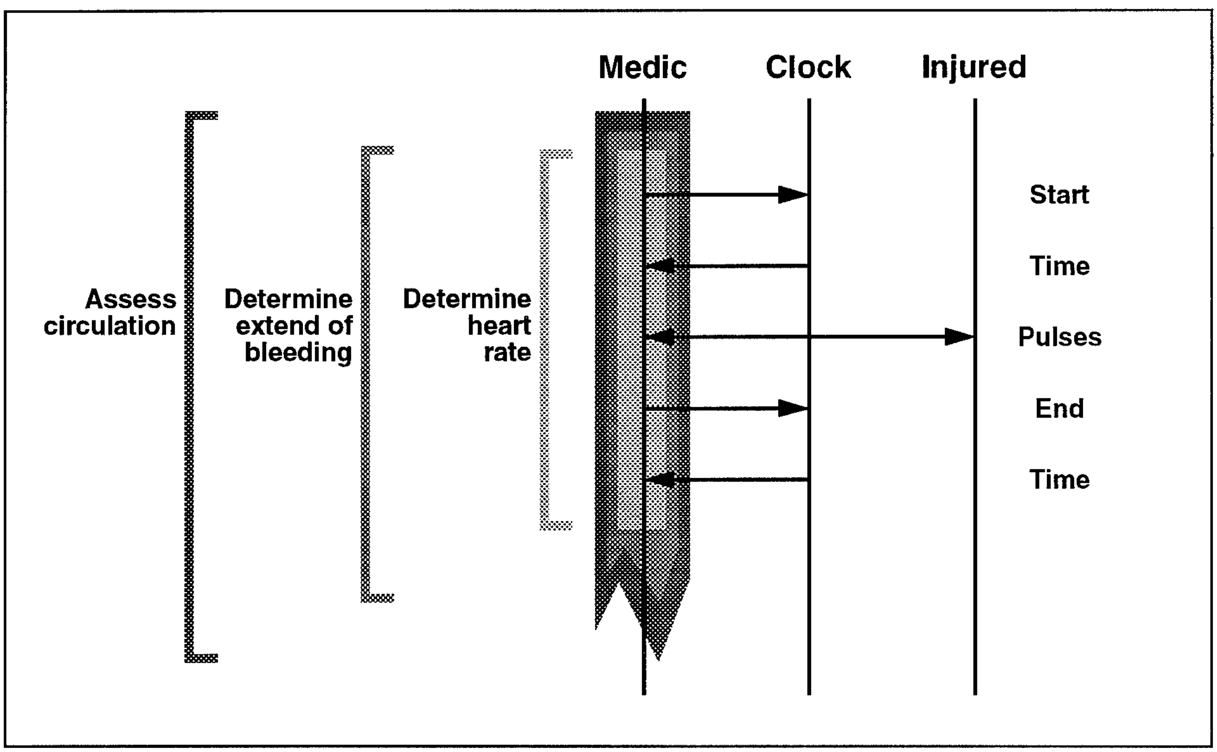

Exhibit 12. Event Trace Diagram

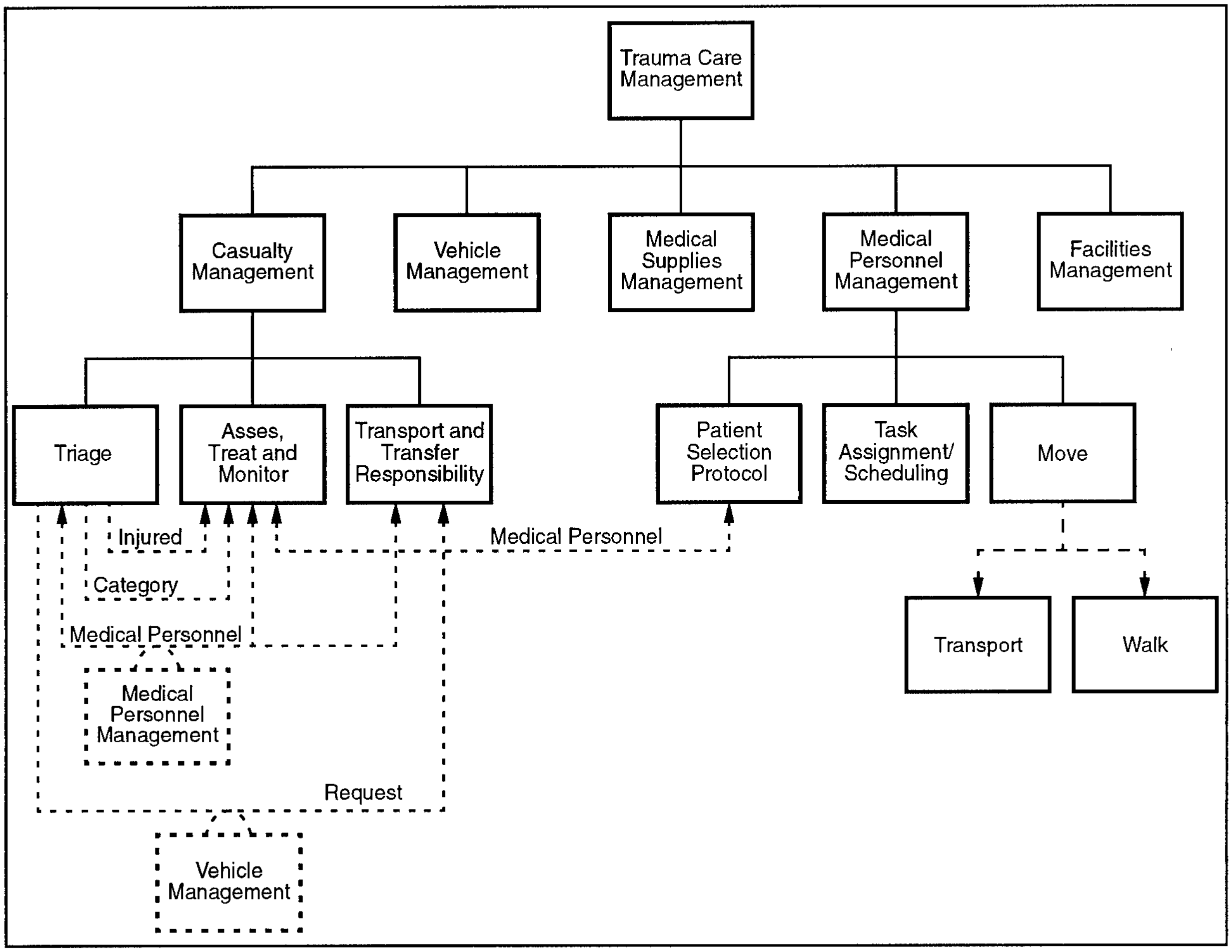

Exhibit 13. Conceptual Hierarchy

5.2.3 Concept Notations

Concept notations are used to differentiate between domain concepts. Concepts identified by an informant can be represented in a variety of ways. If the audience, at any level, is unfamiliar with concepts being represented in other notations, conceptual analysis and representation is vital to a project's success. If the audience is familiar with the domain and its terminology, the need to identify and represent primary concepts is reduced. Providing concept representations is also critical when the audience is diverse, such as a consortium rather than a single company. Identifying and modeling domain concepts allows all participants to understand and agree on common domain terminology.

Graphical notations that can be used to represent concepts include conceptual hierarchies, concept maps, and object diagrams. Exhibit 13 provides an example of a conceptual hierarchy. Conceptual hierarchies show aggregation of concepts. This type of view is effective when the domain is large and concepts must be consistent for presentation to a large audience.

Another means to represent concepts is the concept map, shown in Exhibit 14. This representation is created by having an informant describe the various attributes within a familiar domain. To create the concept map shown in Exhibit 14, the informant, an emergency medical technician, was

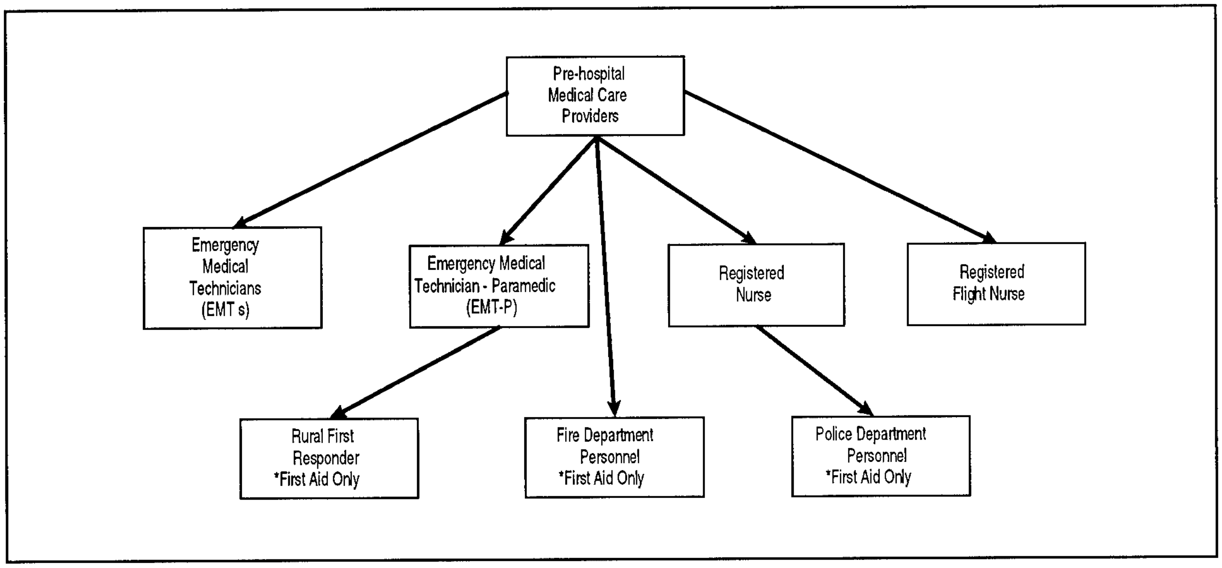

Exhibit 14. Concept Map

asked to describe pre-hospital emergency medical providers. This map was then used to further define attributes and tasks within the overall medical scenario.

A somewhat more formal notation for representing concepts is the object diagram with relations, shown in Exhibit 15. This notation is most appropriate when the audience includes a software development community of practice. Object maps can be used to elaborate a concept map when the concepts involved form a taxonomic classification.

Concept diagrams differ from task diagrams in that they concentrate on declarative, rather than procedural knowledge. They share with task diagrams the ability to represent both semantic and episodic knowledge. Typically, concept diagrams represent static entities, like categories or objects and their properties.

The conceptual hierarchy notation is essentially a variant of the hierarchical process decomposition notation, with the addition of coordinating processes and the introduction of dependencies. The basic notation can be used to represent variability. For example, in Exhibit 13 "transport" and "walk" are two alternative ways to "move" medical personnel. The concept map notation is informal and imposes very few constraints on its use. Thus the variability here is broad, extending to practically any imaginable set of relations among a set of concepts. One can think of the concept map as a precursor to a formal entity-relation (ER) or object diagram. In object diagrams, relations defined among objects can be used to infer variability, based on the constraints that may be inferred from these relations.

5.2.4 Taxonomy Notations

A taxonomy relates concepts in a hierarchy, where each relationship is *specialization* (i.e., more specific concepts appear below more general concepts). There are a number of notations that can be used to represent taxonomies. Two example notations for taxonomies are shown in Exhibit 16.

Taxonomies are hierarchical, as are parts breakdowns, organization charts, status relations (pecking order), training/education credentials (specialists, etc.). Almost every field of study uses taxonomies; in fact, many of the notations we have seen so far in this section can be seen as special

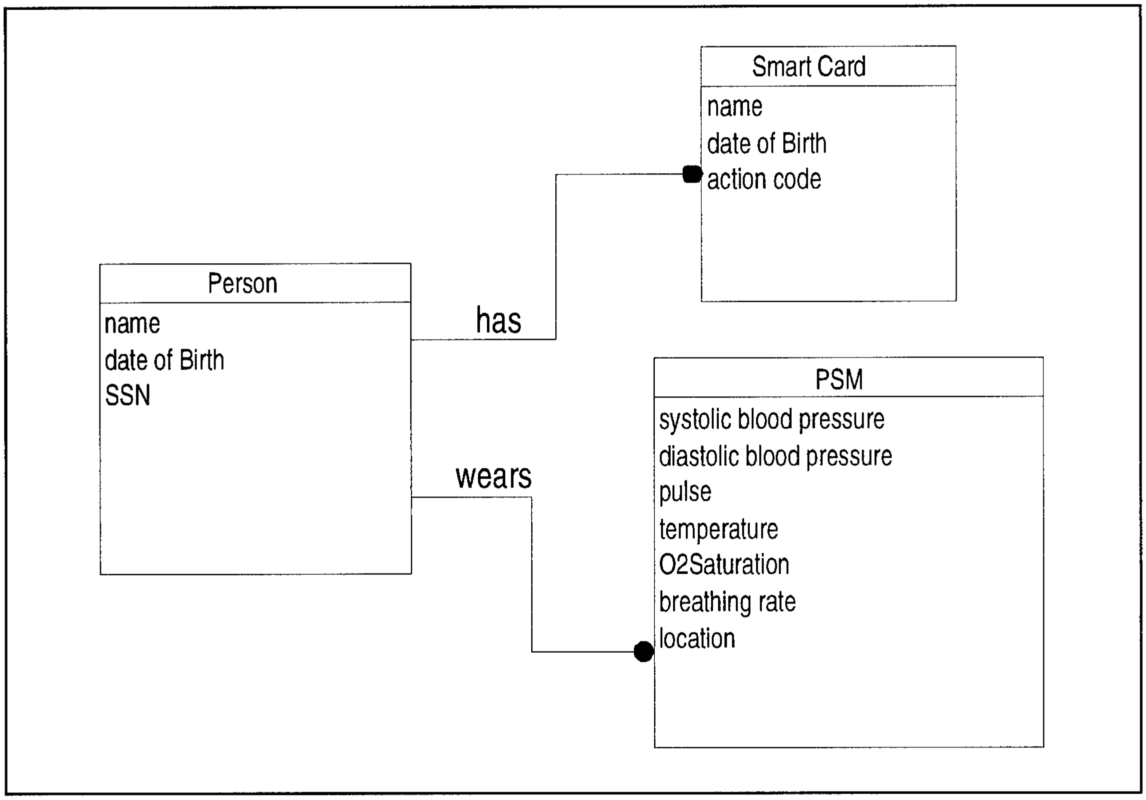

Exhibit 15. Object Diagram with Relations

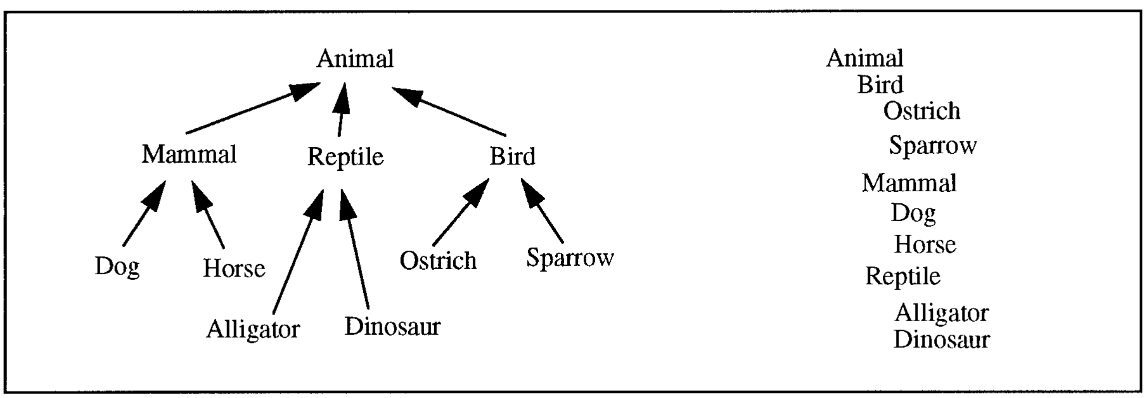

Exhibit 16. Two Ways of Representing the Same Taxonomy, as a Tree and as an Outline

cases of taxonomies. One might think that this ubiquity would make taxonomic notations unproblematic for use in knowledge representation. However, this familiarity of hierarchies can lead to problems in taxonomy interpretation by informants, since the semantics behind a specialization taxonomy are fairly abstract. In a given knowledge acquisition session, if the informant has a high degree of interest in one kind of hierarchical relationship and is shown a taxonomy (particularly in outline form), the informant may conflate the semantics of the taxonomy with the semantics of their hierarchy of primary concern. For example, if a medical practitioner sees the concept "sur-

geon" under the concept "physician" (meaning, according to the semantics of the taxonomy, that a surgeon is a special case of physician), she might misinterpret this to think it says that a surgeon is less qualified than a physician, or that surgeons report to physicians. The warning to take from this when planning knowledge acquisition is that it is important to understand the interpretation placed on hierarchies by the intended audience of a taxonomic representation. It is always the case that the choice of a notation will constrain the choices for a workproduct audience; in the case of taxonomies, the ubiquity of hierarchical diagrams makes this problem particularly subtle.

Taxonomies, like concept diagrams, typically are used to represent declarative knowledge. Taxonomies can represent static relationships between entities, and are not particularly suited to representation of dynamic information. However, taxonomies do explicitly allow the representation of variability through different specializations of a single concept. When combined with relations among the concepts (as is done by modeling frameworks such as RLF [49], [50]) the details of the variation can be expressed. Variability can be expressed at several levels, indicating further and further refinement. In Appendix B of this guidebook, we demonstrate how a taxonomic modeling method can be used as a notation for representing the variability in a knowledge acquisition plan itself.

5.3 Guidelines for Selecting Notations

In this section, we presented four types of notations, as examples of the range of choices available for representing knowledge. Depending on the notations that are familiar to investigators working on a particular project, notations other than those suggested here can be used. Each notation brings with it certain capabilities, advantages, and disadvantages. The choice of which notation to use in any particular knowledge acquisition session will depend on a number of project-specific factors, including the particular knowledge to be captured, the audience for the knowledge, and the skills of the investigators at using and applying a particular notation.

The fundamental attributes presented in Section 5.1 can be used as criteria for selecting notations. Notations can be characterized in terms of these fundamental attributes. Trade-off analysis can be performed based on which characteristics are important to the knowledge acquisition session.

As an example of one of the trade-offs involved, consider a project interested in capturing variability and dynamic information. There is a fundamental trade-off between how variability is represented in dynamic and static notations. At one extreme we find notations that keep tight control over explicitly representing variability (e.g., taxonomies), while at the other, we find notations that represent variability only implicitly or not at all (e.g., scenarios). The trade-off comes when we compare the nature of the things that can be modeled by the notations; taxonomies tend to be static descriptions of terms or concepts, while scenarios model dynamic processes and interactions. The more dynamic the target of the notation, the more difficult it is to represent variability explicitly.

If there are multiple audiences interested in the results of a knowledge acquisition effort it is possible that they will have incompatible representation needs. Specifically, the focus community may find value in having a reference source, but the formal notations favored by the technologists will be of little use to them. It may be difficult to find a single compatible notation that will serve the needs of multiple audiences. On the other hand, use of multiple representations presents other problems. In the absence of appropriate automation, maintenance of multiple representations of the same information presents configuration management problems and labor-intensive project scenarios. For information that is essentially static this is at least not a recurring cost; but some changes are almost always necessary, if only as part of validation. In any case, it will be necessary that some individuals be on staff who can translate among the various representations.

The choice of a notation to use for representing knowledge will always bring some bias with it. For every notation, there are some things that are easier to represent, and some that are more difficult. This can be a blessing or a curse. On the one hand, the biases of the representation can interfere with the acquisition of information from the informant. On the other hand, the biases inherent in a well-defined notation can be worked out in advance, and the notation can be chosen for situations where it is deemed that its bias presents little danger to the project. Thus biases brought to a session by representing knowledge in chosen notations can be easier to manage than human biases, since human biases may change and cannot always be known in advance.

6.0 Dossier Planning and Management

Management of the ***dossier***, the repository where all materials produced in the knowledge acquisition effort are stored, is central to managing the overall knowledge acquisition effort. The dossier contents can include the original workproducts from the focus setting, (manuals, textbooks, legacy code, screen dumps, demo materials, program documentation etc.), as well as derivative workproducts produced by the knowledge acquisition effort itself. These could be videotapes of interviews, reports based on notes taken by the investigator during an interview, high-level diagrams such as those outlined in Section 5.0, lab notebooks (raw data) from experiments etc. Although the dossier could very well include a large amount of off-line material, its index could still be supported by a on-line indexing mechanism.

The dossier plays several critical roles, including the following:

- *Serves as the delivery vehicle to the project audience.* First and foremost, the dossier is where the workproducts that are readable by the audience of the project will be stored. According to Canvas, every workproduct has a specific audience for whom it is intended; the value of the project hinges on how well the project audience can find and understand the workproducts that are intended for it.
- *Directly records the artifact threads.* Of the three types of thread that make up the fabric of the knowledge acquisition project, two of them, the informant and investigator threads, have to do with the state of mind of human beings, and hence cannot be stored directly. However, the third thread, the artifact thread, consists (as described in Section 3.2.3) of a series of annotations to a workproduct. Each of these annotations will be stored in the repository; a well-structured repository will show the heritage along this thread, allowing future plans to reference it, to determine what has already been done, under what circumstances, etc.
- *Provides source materials for improving investigators' knowledge.* A project with a large number of investigators will have to manage the gaining of knowledge by these investigators carefully. As the project progresses, more and more knowledge specific to the KA project will be available, and the informants who are most deeply involved in the project will come to expect that this information can be used as a base for further discussions. The well-structured dossier will also serve as a source of training material, where new investigators can catch up to become productive, or investigators with longer term commitment to the project can keep abreast of one another.
- *The dossier index should reflect planning criteria.* In order to place a particular session into the knowledge acquisition planning canvas, it is necessary to know a number of things about its participants and goals, e.g.,

  — "What role does this informant play in his work setting?",

  — "What sort of knowledge (task, concept, scenario) is needed?",

  — "In what topic are we interested?", and

  — "For which audience is the output intended?"

  All of this information, needed in planning, should also be stored as index material in the dossier. The data dictionary for this index then serves as a guide for planning.

In Section 6.1, we will provide detailed guidance for creating an index for a dossier. In Section 6.2, we will show some sample scenarios for how to use a dossier index that was created

in that way, and finally in Section 6.3, we discuss possibilities for automated support for searching the index to support those scenarios.

6.1 Structuring the Dossier

A great deal of the basic dossier structure can be laid out during up-front planning. In this section, we will show three "starter sets" for the top levels of three different hierarchical organizations for the dossier, based on three key elements of session planning. The intent of the "starter sets" in this section is to cover the main elements identified in Canvas, so that a dossier structuring effort need not begin with a clean slate each time. In any particular project, these starter sets should be modified; some categories suggested here might not be applicable to the project, others might need to be specialized into more detailed categories, and/or completely new categories might need to be added. Along with each starter set, we will provide a set of questions that can help in extending or restricting the set. Sometimes the questions will be yes/no questions that determine whether a particular category should be included. Others will be more open-ended questions, designed to elicit further categories.

We will provide starter sets for the following three categories:

- the intended audience;
- the knowledge source; and
- the selected knowledge representation.

These categories are particularly useful because a great deal of information is known about them at planning time, allowing them to be used for the initial structure of the dossier. In any planning effort, more categories might turn out to be useful for indexing. In this sense, the list above is also a "starter set."

As part of the TCIMS knowledge acquisition effort, a large dossier of derivative knowledge acquisition work products was created under the name SEPWeb. After each starter set, we will illustrate how it can be refined to form a working index of the materials in SEPWeb. Since the SEPWeb is still in development, we do not attempt to make a report of the state of the search capabilities of SEPWeb at any one point. The diagrams in this section are indices based on the publicly available data in SEPWeb, and were constructed as illustrations of the principles outlined above. The examples here have been simplified from the SEPWeb and TCIMS material, both for illustrative purposes, and to protect TCIMS consortium confidential information.

The hierarchies created from these starter sets can be used as indices by attaching actual dossier entries at the leaves; thus, all documents intended for "physicians" as an audience will be collected in the "physician" node in the audience hierarchy; all documents using representation notation "Object diagram" will appear under that node in the representations hierarchy. In this way, each hierarchy provides a different viewpoint on how to organize the corpus of material in the dossier.

6.1.1 Audience

Exhibit 17 shows a starter set for an index of audience types. The top level is taken from the overall model of the knowledge acquisition session that has been used throughout Canvas.

In order to customize this model for your project, the following set of questions might be useful:

- Is the goal of the project to provide new technology for the focus community? If so, then there are probably one or more implementer audiences in the project, as well as one or more user communities in the focus audience.
- Is the goal of the project to change work practice in the focus community? If so, then some group of policy makers will be an audience for the project.
- Is the goal of the project to return information to the focus community? If so, then a student community will be an audience for the project.

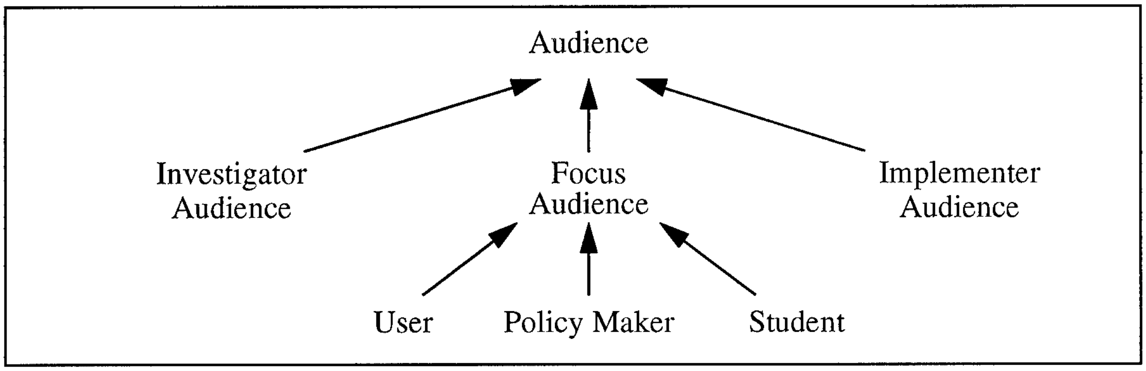

Exhibit 17. Starter Set based on Intended Audiences

When tailoring Exhibit 17 to a particular project, the focus audience will probably be the section of the diagram that is decomposed the most; typically this audience will consist of a large number of distinct communities of practice, with different conventions, notations, and expectations. An example of this is the hospital environment, where doctors and nurses are two finer-grained subgroups of the focus community. Documents intended for use by doctors will be stored under the "doctors" node, while documents intended for nurses can be stored under the "nurses" node. Documents intended for the more general medical audience would be stored under some higher level common parent to these two nodes.

The starter set shown here is intended to be suggestive of where boundaries between different audience groups are often found. The term "Users" in Exhibit 17 refers to practitioners in the focus community who perform activities that make use of systems built from the knowledge acquired in the project. These could include virtually any kind of practitioners (e.g., secretaries, accountants, researchers, developers, etc.). There will often be several user groups, possibly with further specialization relations. Policy makers refer to people who make decisions about the work setting: directors, technical advisors, and project managers are typical examples. The set of indices based on distinctions made in the focus audience is likely to develop as the project proceeds, and new categories of practitioners are identified, or new distinctions of work setting are discovered among known groups.

> Example. Exhibit 18 shows the index of audience communities in SEPWeb. In this case, the greatest variety of communities was found within the focus community. Following the set of questions above, we find that one goal of the project is to change the work practice in the medical and military communities; therefore, these two communities appear in the index. The questions also indicate that policy makers should be included. In this case the policy makers are the people who perform administration of the clinics. Within this group, we have another level of distinction between the clinic directors and the receptionists who keep track of the everyday business. Since we are interested in returning information to the medical part of the focus community, medical students are also included.

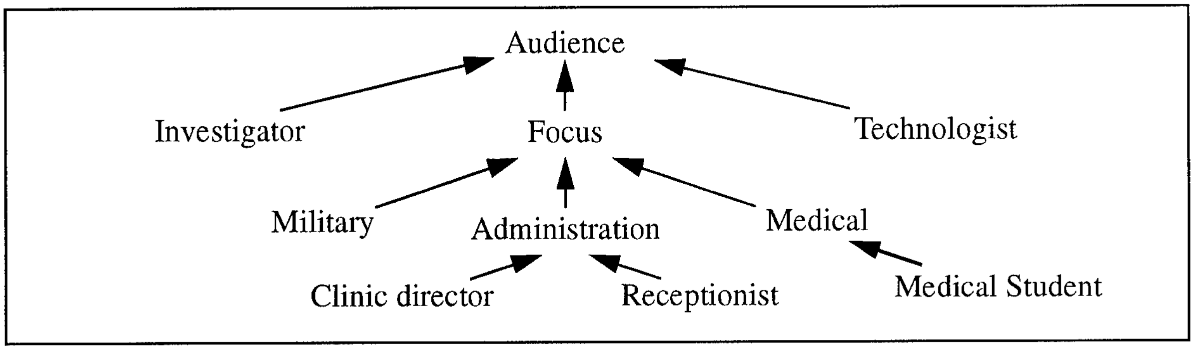

Exhibit 18. SEPWeb Index based on Audiences

6.1.2 Knowledge Source

Exhibit 19 shows a starter set for an index based on knowledge sources. The two main categories of knowledge source, informant and artifact, form the basis of the index. To refine these categories further, the guiding principle is to consider types of knowledge source that are likely to shed a different viewpoint on some topic, and hence could be used to help someone find the resulting knowledge acquisition workproduct, when he shares that viewpoint.

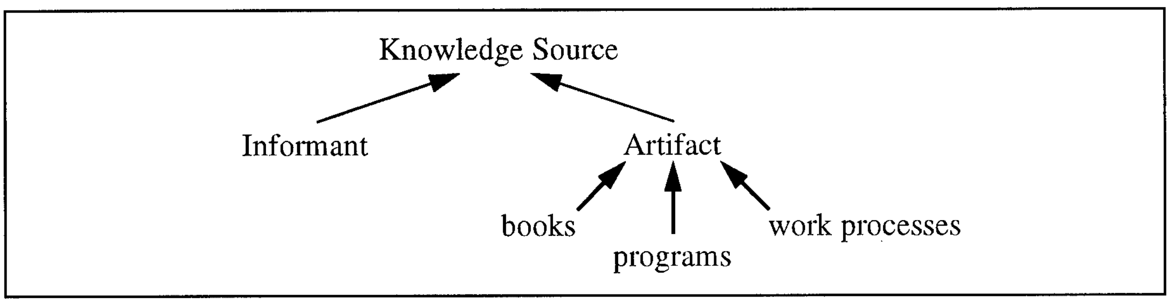

Exhibit 19. Starter Set based on Knowledge Sources

For informants, the following set of questions can help to find such categories:

- Given an informant from one community of practice, with what other communities do its practitioners commonly interact? Builders and electricians are an example of communities of practice that relate in their professional activity.
- What are some well-known conflicts between communities? Copy editors and authors are an example of this sort of division.
- Is there professional stratification within the community? Doctors and nurses are an example of communities separated by professional class.

This starter set is quite open ended, and can be extended with an elaborate structure, depending on the details of the focus community. The first few steps of such an extension are shown in the example below.

> Example. Exhibit 20 shows the index of knowledge sources for the workproducts in the SEPWeb dossier. The starter set shown in Exhibit 19 suggests that we should consider the artifacts that were studied as well as the informants. However, since TCIMS was primarily an

interview-based knowledge acquisition effort, there are few formal reports based upon artifact studies. On the other hand, several distinctions among informants were made, only a few of which are shown here. The first distinction, between "medical" and "transport," exemplifies distinct communities that interact professionally in the TCIMS domain; in particular, medical personnel often have to be transported to accident sites, and to be transported with patients safely back to a medical facility. Transportation personnel have to be able to interact with the medical personnel in order to facilitate this transfer.

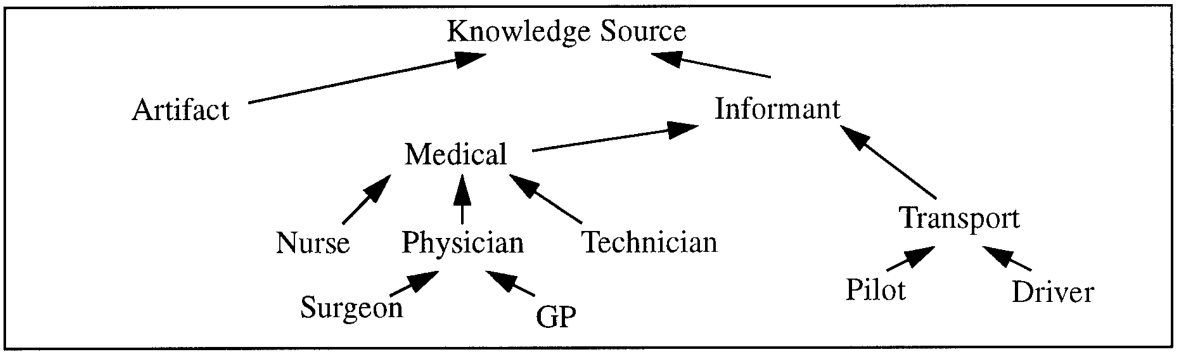

Exhibit 20. SEPWeb Index based on Knowledge Sources

Within the medical community, the professional stratification of physicians and nurses is recorded in the separation of nodes for those two groups. From these two groups, following the questions above, we find the community of lab technicians, with whom both of these commonly interact. Finally, surgeons and general practitioners are separated within physicians, because of the difference in viewpoint that these two groups often have.

6.1.3 Knowledge Representation

Exhibit 21 shows a starter set for an index hierarchy of the major types of notations used for the representation of knowledge, as described in Section 5.2. For each of these, a number of particular representations are available. We will not provide a complete catalog of all such notations from the literature. Depending on the choice of particular notations, there might be further refinements of each of these types.

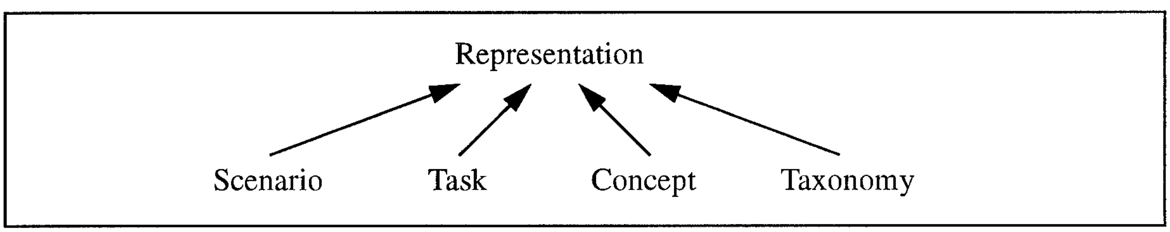

Exhibit 21. Starter Set based on Knowledge Representations

The knowledge representation choices identified in this hierarchy should be based on the decisions made during enterprise planning, when the representations were matched to the audiences (Section 4.1.4). The attributes mentioned in Section 5.1 can be used to select representations; they are summarized in the set of questions below:

- Will procedural or dynamic knowledge be collected during the project? Procedural and

dynamic knowledge is usually best represented using scenario or task representation notations.

- Will there be specific case study information for procedures? If so, it can be represented using scenario notations.
- Is the project interested in hierarchical classification of information? If so, the hierarchical classification can be represented using taxonomy notations.
- Does the focus community use taxonomic notations already? If so, then the dossier should be organized in such a way that the different taxonomic notations will be clearly distinguished.
- Will there be static knowledge of the attributes and relations among entities? If so, then the static knowledge can be represented using concept notations.

This set of questions should be extended, along with the starter set, to accommodate the representation notations that are familiar to a particular knowledge acquisition team. The following example shows how this index has been expanded in the case of SEPWeb.

> Example. Exhibit 22 shows the categories of representation notations used in the SEPWeb. Each dossier entry is labelled with the notation used to express it. Following the set of questions given above, since TCIMS concentrates on acquiring process information, both in the case study form (scenarios) and general form (tasks), these two notations appear in the SEPWeb index. In TCIMS terminology, "object representation" is the name given to a taxonomic model. TCIMS is a large, comprehensive effort, so it includes all the types in the starter set, along with a further type, the flow representation, for recording coordination information among the many actors in a military medical situation.

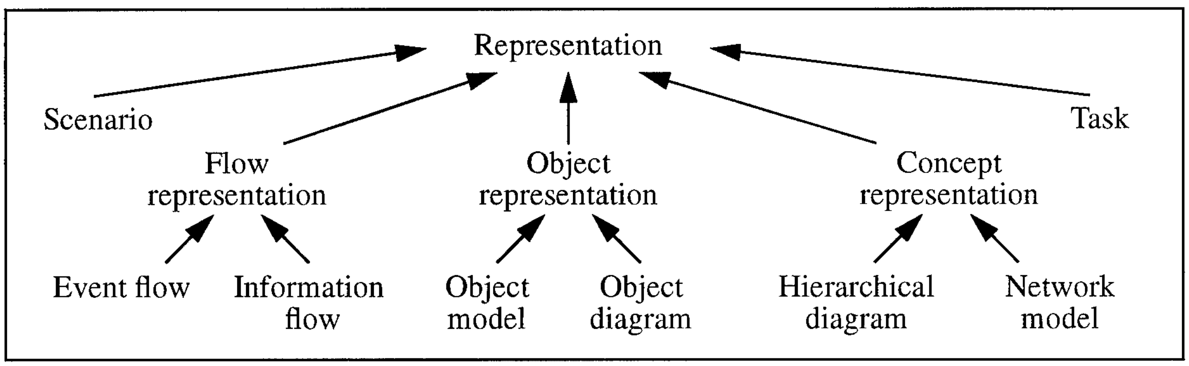

Exhibit 22. SEPWeb Index based on Knowledge Representations

6.1.4 Topic

One of the most powerful methods for indexing a dossier is by topic. Recall that in Canvas, the word ***topic*** refers particularly to something known to the focus community that is the focus of attention in some KA session. No matter who is consulting the dossier, it is likely that they will have some idea of what topic they are interested in learning about. A particularly useful option for managing a dossier based on topic is to use keyword searches, since the keyword lists can easily be updated along with the knowledge acquisition workproducts themselves. In a large dossier like SEPWeb, this can lead to problems when the set of keywords becomes too large. SEPWeb solves this problem by structuring the set of keywords into three sets, one each for the scenario representation, the task representations, and the other (flow, object, concept) representations. Further

structuring of the keywords is possible, and is the subject of further research for the SEPWeb team.

6.2 Sample Usage Scenarios

The structure of the dossier can be used in many ways to support processes involving the use of the dossier. These range from planning the knowledge acquisition process itself, to access by end users of the dossier. We will outline a number of scenarios, and show how they make use of the indices outlined above.

The scenarios given here are only the tip of the iceberg. By including a few example queries from different possible audiences including the knowledge acquisition team itself, the intended audience, and other audiences, we have tried to give a flavor of the role a dossier structured according to the Canvas framework could play in ensuring a knowledge acquisition project's success in meeting objectives, and obtaining maximum value from its results.

6.2.1 Use in Managing the Ongoing KA Effort

The elements in the dossier can be used during the knowledge acquisition effort to manage all of the threads. Some example uses:

- *Perform a "walk-through" of a KA workproduct with an new expert.* (i.e., not the original source of the knowledge in the KA workproduct). To acquire more detailed knowledge, an investigator can show one of the KA workproducts to another expert, and elicit commentary. To do this, he needs to use the audience index to verify that the new expert is in the intended audience, or to handle the resulting difficulties if he is not. This usage corresponds to evolving the project along an artifact thread.

  Example. We might want to ask nurses to review reports of procedures given by physicians, to see what their viewpoint can bring to the picture. We can use the audience, knowledge source, and representation indices to support this plan; we use the knowledge source index to determine if the workproduct came from an interaction with a physician (including surgeons and general practitioners), we use the representation index to restrict it to task representations, and finally, the audience index to make sure that we are using a diagram that was intended for a medical audience, so that we can reasonably expect the nurses to understand it.

- *Perform a comparison or correlation analysis.* We might want to study several knowledge acquisition workproducts at once to find correlations or comparisons. We can use all the indices, depending on what we want as the common theme of the comparison.

  Example. A systematic comparison of nurses' and physicians' views on process could yield an interaction diagram that shows where the two communities have to negotiate some limited resource. Such a comparison would make extensive use of the knowledge source index.

- *Have a new investigator interview an informant who has been interviewed before.* If a new interviewer wants to follow up another interview, he might want to read up on the reports that were elicited from this informant before. This usage corresponds to tracking the evolution of the project along an informant thread.

  Example. A physician might have modified his view on a particular procedure after seeing the comparison between the nurses' and physicians' viewpoints. Investigators pursuing further interviews with this physician should be aware of this history. The knowledge source index can be used to find the relevant workproducts.

- *Prepare an investigator to interview a new informant.* If an investigator wants to interview a new informant, she might want to "do her homework" so that she can maximize the benefit of the time with the informant. This could be achieved by examining other knowledge acquisition workproducts that are already available from sessions with other related knowledge sources. This usage corresponds to evolving the project along an investigator thread.

  Example. Suppose that a large KA project gains access to a high-profile expert to use as an informant. She is the only known source of information about the effects of residual radiation on military installations; however, in order to understand her theory, one needs to know a number of details of how military installations deal with hazardous materials. The planner decided to have the interviewer prepare by examining all the relevant workproducts in the dossier.

6.2.2 Intended Use by Target Audience

The dossier could have a number of intended audiences, managed by the audience index described above. When the dossier is used by a member of one of these audiences, it can be used to do the following:

- *Gain an overall view of some topic.* In this case, the user would want to find all workproducts, of any representation type, intended for any audience of which he considers himself a member.

  Example. If a clinical director wants to have the big picture about blood supplies, he can use the audience index to filter the workproducts to find those that are intended for himself as an audience, and use the topic keyword list to find the workproducts relating to blood supplies.

- *Find out what practitioners in some setting say about some topic.* The user can consult the sources index to find the workproducts that were produced by some other practitioner group.

  Example. A physician is interested in knowing what lab technicians have to say about data confidentiality. He can filter using the knowledge sources index to find only those workproducts.

- *Locate a specific report.* All the indices can be used together to find a particular report.

  Example. A system implementer is planning the specifications of a new program to be used by the BAS surgeon. She wants to know all the interactions that the BAS surgeon has with other agents. She begins with the BAS surgeon from the keyword list; this results in several workproducts. Since she knows that she is interested in relationships between entities, she can restrict her search to concept representations. She finally finds the workproduct she needs under the network model of the BAS surgeon.

6.2.3 Future Spin-off Uses

In some sense it is contradictory to plan a dossier for unplanned uses. However, if we track the context of each knowledge acquisition workproduct, then there is at least enough information to determine what parts of the dossier can be useful.

  Example. A medical student working with one of the physicians interviewed in the TCIMS project decides that he would like to use the dossier as source material for a class. He can use the audience index to determine which knowledge acquisition workproducts were intended for a medical audience (since those intended for other audiences might have inaccuracies or imprecisions that are of concern only to medical practitioners). Similarly, the knowledge

source index can help him to determine the reliability of the particular information.

6.3 Possibilities for Automation

The usage scenarios given above suggest ways in which the dossier could be indexed, so that knowledge acquisition workproducts can be stored and retrieved in useful ways. In order to gather the information needed to form these indices, the knowledge acquisition sessions must keep careful track of the information used in workproducts, in particular, the source of the information, the intended audience of the write-up, the representation that is used, etc. The exact details of which information will be included in the indices should be determined during planning of the enterprise, as described in Section 4.1. The KA workproducts must then be stored according to these indices, and made available to the entire investigator team. Given that this team might well be widely distributed geographically, this distribution is a perfect opportunity to use the capabilities that are provided by the technology of the Internet.

The wide variety of types of information available to index the dossier provides a wide range of opportunities for automatic support. In this section, we outline these possibilities, some of which are already under development by the Canvas team.

6.3.1 Web Accessibility

Many of the requirements of a knowledge acquisition dossier are similar to those found on the World Wide Web (WWW), namely, that a large corpus of information, divided into separate pieces, needs to be viewed in a flexible and exploratory way from geographically diverse locations. Entries in the dossier are related to one another in many different ways. This suggests that we might use some of the technology already available on the WWW to automate the dossier.

One capability that has already been tested with SEPWeb is to provide a search engine for a set of dossier entries, which can find an entry based upon keywords. If we associate keywords with topics, this can be an effective way to find a workproduct related to a given topic. However, often a user sees a topic mentioned in a workproduct, and would like more information about that topic. SEPWeb already implements hypertext links for such references in text-based workproducts. For diagrammatic workproducts, the user could return to the search engine and search for the new topic, but far better would be a hypertext link from the original workproduct diagram to a list of workproducts about the related topic. The problem with this approach is that it involves constructing a hypertext active display for each workproduct, and linking it to its related topics. For general diagrams, this could incur considerable expense.

This expense can be reduced by providing a set of standard representational notations, along with web-aware tools for representing knowledge using these notations. One drawback of such an approach is that it limits the range of notations that can be use. A far more serious drawback is that the notations must be sufficiently formally specified that they can be automatically processed to create the hypertext diagrams. Many representational notations (especially those whose audiences are not technically oriented) get much of their power from their informal flexibility. It is not uncommon to mix natural language with an approximation to a more formal notation (such as entity relationship diagrams or flow charts) in a single representation; requiring that the representations meet some notational standard could impair this flexibility.

6.3.2 Hypertext Organization of Indices

One way to make use of hierarchical indices such as those given in this section is to provide the hierarchical structures in an outline form, complete with the usual outliner functionality of collapsing subtrees, displaying to a particular depth, etc. Such a capability would support usage scenarios where the user is not certain in which category she is interested. If, for example, she was interested in emergency room practices, she could, by browsing under the node for medical personnel, find the nodes for nurses and lab technicians. Once a node has been found, then all of the workproducts that are sorted under that node (e.g., all the workproducts for which nurses are the intended audience) can be accessed directly.

6.3.3 The Reuse Library Framework

Although all the capabilities described above are available in one form or another, there does not yet exist a comprehensive tool for supporting knowledge acquisition in Canvas. One candidate that provides much of the necessary support is the STARS Reuse Library Framework (RLF) [49]. RLF is a framework for creating hierarchical indices of libraries of complex objects. It was designed, as its name suggests, as an infrastructure for managing libraries of reusable software components. RLF brings with it a semantics for hierarchical models of classes and instances; Appendix B gives details of the hierarchical modeling language behind RLF, as well as examples of modeling the knowledge acquisition process itself in RLF. For the purposes of Canvas, RLF classes correspond to the nodes in the hierarchies given in Section 6.1. The instances of each class correspond to the KA workproducts themselves. Tools for navigating RLF models over the WWW have been prototyped, and the models of the SEPWeb corpus have been tested. Preliminary results suggest that these structure, represented in RLF and viewed with its web-based browser, give the user a quantitatively different view of the materials than the keyword search alone.

6.3.4 An Automated Scenario

All of these capabilities can best be illustrated through a sample scenario that takes advantage of all the automated capabilities discussed above. In particular, the keyword search capabilities of the web can be combined with the hierarchical structured indices to exert fine control over the dossier.

> Example. Suppose that a system developer is interested in how the duties of a midwife are carried out in a modern hospital. He is interested in knowing which personnel are responsible for which duties, and what is the sequence of back-up support in the event of complications.
>
> He begins by using the keyword filter on words such as 'obstetrics,' 'childbirth,' and 'maternity.' This filters the dossier down considerably, but there are still a large number of workproducts available. He begins by looking in the knowledge source hierarchical index under the Medical node in Exhibit 20 and finds that there are still too many workproducts to examine sequentially. So he descends toward Physician, and finally to Surgeon, where he finds two workproducts, one a task analysis of the Caesarean section procedure, and the other a concept diagram of drugs for cervical dilation. A search under Nurse instead again shows too many workproducts; however, a further keyword filter on "Caesarean" brings up a task analysis for preparing the patient for the procedure, as well as a comment on "breech delivery". A further search on breech delivery under Nurse shows task analyses for procedures to turn breech babies before delivery.

Navigation through the dossier can continue in this way, alternately using keywords and hierarchical structures to limit the number of workproducts accessed. If all capabilities are available on the WWW, then this capacity is accessible to members of the target audience all over the world. The WWW also provides the capability to view the KA workproducts themselves once the search has been sufficiently limited.

6.3.5 Future Work

RLF already has automated support for many of the capabilities in the scenario above, including an outliner that works on the WWW. This means that a user can examine the structure of the dossier index, and view a page from the dossier itself as needed. Further work on RLF includes support for model types other than hierarchies, and a wider array of ways to view and navigate through multiple hierarchies.

7.0 Conclusions

In this guidebook, we have presented an approach to organizing a knowledge acquisition project that includes insights gained from the point of view of the Organization Domain Modeling (ODM) and Scenario-based Engineering Process (SEP) methods. Many of the insights were gained from review of a very large knowledge acquisition effort, TCIMS, which has been a source of examples throughout the guidebook.

The guidebook has described the kinds of activities that can usefully be considered as knowledge acquisition, as distinguished from many closely related forms of learning and knowledge transfer. A conceptual framework has been offered that clarifies the role of various communities of practice (focus, investigator and target communities) and basic elements of the knowledge acquisition process (i.e., the role of investigators, informants, and artifacts in acquiring knowledge about given topics within specific settings). These are integrated in an overall metaphor of the knowledge acquisition "canvas" that includes multiple threads of interactions among the various elements. The guidebook has also presented specific recommendations for planning and managing a knowledge acquisition enterprise in light of these basic concepts, as well as implications for automated support of aspects of the knowledge acquisition process.

Throughout this guidebook, certain key principles have emerged as recurring themes in both the concepts and specific guidelines offered. This section presents our conclusions by describing some of these key Canvas principles, and proposing some areas for future research.

7.1 Canvas Key Principles

This section summarizes and recapitulates six *key principles* that describe what it means to do knowledge acquisition "according to Canvas," and links them to their implications in the planning of a KA effort. These principles reflect "best practice" in the KA field. For some readers, these principles may seem to be obvious statements of common sense, too plain to merit mention. We hope these principles will help other knowledge acquisition practitioners to reap the full benefits of their efforts, even on KA projects that are smaller than TCIMS. Experience demonstrates that considerable skill and care is required to apply these principles systematically on a knowledge acquisition project. To other readers, these principles might challenge long-held assumptions that have been considered incontrovertible or self-evident. We believe they are important principles to acknowledge in order to make consistent and dependable use of knowledge acquisition results, and to better understand the potential value of those results.

- Bias is unavoidable, but its effects can be managed.

  As human beings try to make sense of their world, they take advantage of patterns found in their experience. While this makes the flood of information in the world intelligible, it also means that one's interpretation of the world depends on one's own background. This influence is known as *bias*.

  Through systematic tracking of exposure to information, bias in a knowledge acquisition project can be controlled. This requires careful planning and record keeping of the activities of each actor in the knowledge acquisition team.

- Variability is an integral part of knowledge.

  People will disagree on almost any topic. Some disagreements are the result of one person being misinformed, but the most interesting disagreement comes from difference in background, interests, or viewpoint.

It is easy to adopt strategies that minimize variability in KA, but in so doing we may lose a rich source of data. Differences need not be ironed over in collecting a repository of knowledge; the differences themselves are part of the rich, cultural web of the knowledge. For some purposes, such as domain engineering, variability is an essential aspect of the data to be gathered. A number of representational notations can express this variability, leaving the audience to decide which variant is appropriate to any situation.

- Cultural differences between communities of practice are the greatest barrier to knowledge transmission.

  If people from varying backgrounds could communicate detailed technical information easily, there would be no need for knowledge acquisition. In order to make sense of a complex world, people rely on simplifying assumptions that are often difficult to articulate. Cultures grow around shared sets of these assumptions. Some of the techniques for knowledge acquisition explored in this guidebook are adapted from social science research, where "culture" in this sense may be interpreted quite broadly; but the techniques can also be adapted to the "micro-cultures" of particular organizations, technology development environments and work settings.

  Participants in a knowledge acquisition project play an "ambassador" role in making cultural assumptions explicit so that information can more readily flow across the boundaries of different communities. Awareness of the distinct communities involved, and care in mapping terminology, etc. from one community to another, are key aspects of a systematic approach to knowledge acquisition.

- Anyone can be a source of knowledge.

  So-called "experts" are not the only people who have knowledge; everyone who accomplishes some task exhibits knowledge. No community or practitioner has a monopoly on "important" knowledge.

  A comprehensive picture of an interaction among communities of practice will include information from all of the communities involved, and a broad selection of practitioners in each of those communities.

- Eliciting knowledge will intervene in the studied setting.

  The knowledge acquisition process will cause people to think about their work in a different way, and to make new connections to other workers. The mere act of reflecting on their work or meeting with other people will cause change in the focus community.

  The knowledge acquisition practitioner should use this as a boon, as a way to create value from the acquisition process itself. Intervention, properly controlled, can be an instrument for improving work practice.

- Representation is key to managing knowledge acquisition.

  The knowledge acquisition process is not complete until the knowledge is written down. This process of distillation and translation can present great challenges for the investigator. Representations should be chosen carefully based on the features of knowledge listed above. Representations must be understandable in the culture where they will be used, and by the individuals who will verify their content. Representations have inherent bias. Some can accommodate variability, while others cannot. The representations are a record that can have value in the studied setting.

Knowledge acquisition is, in some sense, the fundamental intellectual activity of a social creature, gaining knowledge from its fellows. The process of acquiring knowledge has a value in itself that usually goes beyond the value that was initially anticipated. It is surprising that, in many projects, the value of a knowledge acquisition effort is not recognized or appreciated outside the knowledge acquisition team. Knowledge is a powerful and flexible commodity, and it takes creativity to take full advantage of it. A knowledge acquisition team should always be ready to find new and unexpected ways of making its project successful.

7.2 Future Research

We conclude by giving a few ideas of directions for further research that would aid the knowledge acquisition process. Some other ideas along these lines, particularly focused on supporting technology development, have already been discussed in Section 6.3.

7.2.1 Presenting Knowledge to Various Audiences

Given the central role that representations play in the knowledge acquisition process, much future work that we propose in this area focuses on presentation and translation of representations. Representations of knowledge in Canvas are typically produced by an investigator, but they must be read by other people, including the informants from which the knowledge is elicited and the target audience for whom the knowledge is intended. This means that there is a need for careful study of how representations can be presented to various audiences in such a way that the audiences will not misconstrue the representations' content.

A clear example of this problem is the situation with taxonomies mentioned in Section 5.2.4; many communities use taxonomic-style diagrams for many purposes. If we want to present a taxonomy to someone in one of these communities, we need to make certain that we do not collide with any use of similar diagrams with which practitioners are already familiar. This requires a careful study of how people understand representations, including analysis of psychological as well as cultural aspects. Furthermore, some notations have a tight semantics, for which technical training is needed in order to understand the semantics. Some communities might be willing to make the investment to undergo this training (in much the way that psychologists have adopted the study of statistics into their practice), while others might not. At the very least, notations with clear semantic descriptions, as well as automated support, could form the basis for such adoption by a community of practice.

Another issue when presenting a representation to some audience is the cognitive load that the representation places on its viewer. Again using taxonomies as an example, simple trees can be presented a few dozen nodes at a time using presentations such as outliners and tree browsers. Interactive support for such structures allows a user to manage several dozen such objects. On the other hand, if a tree includes more detailed structure (e.g., semantically laden linkages between the nodes, or attributes of the nodes), then the basic outliner and tree browsers lose their effectiveness. The RLF diagrams shown in Appendix B demonstrate the difficulties of showing several interconnected nodes in a semantic net, and a simple solution that was effective in this case.

Future research on these issues would include careful study of how practitioners from different communities understand various representations. The research would also include experimental work in the spirit of psychological investigation about how well subjects can cope with information, as well as cultural studies of the ways in which similar representations are understood in different communities.

7.2.2 Translation Between Representations

One of the constraints on representations in a knowledge acquisition context is that they typically have to serve two masters; first, they have to be understandable to someone who is in a position to verify the correctness of the knowledge represented, and second, they have to be understandable to someone in the target community, which is often more technically oriented. This problem cannot simply be ignored by training one community to understand the notations of the other; many notations are quite technical and require expertise to use properly. Because of the possibility for misunderstanding, poorly formed representations can be more dangerous to cross-community communication than having no representations at all.

One solution to this problem is to let the investigator community play the bridging role by performing translations by hand. The investigators can work with knowledge sources in representations that are familiar to the source community, verifying the information represented. Then they can translate this information into a form that is understandable by the target community. It is then up to the investigators to make certain that the information is faithfully transferred from one representation to another.

Tools could facilitate this endeavor. If all notations involved have a formal semantics, it would be possible to prove whether one representation has the same content as the old. The translation could even be automated; given the semantics of both notations, representations in the source notation could be used to automatically generate representations in the target notation.

A serious problem with this approach is that often the power of a representation lies in the informality of its notation, which allows it to leave certain things unspecified or ambiguous. Translations between informal notations, or from an informal to a formal notation are problematic. Providing tools to support such translations in the context of knowledge acquisition would be a great boon to project planning.

Appendix A: Canvas as an ODM Supporting Method

Although Canvas does not assume a domain engineering context, it directly supports knowledge acquisition as part of the ODM domain engineering life cycle. In theory, Canvas could also be applied to other domain engineering methods and approaches. However, many of the core concepts incorporated in Canvas are directly transferred from their use in ODM, so the approaches are compatible at both the conceptual and process levels. This appendix discusses Canvas as an ODM supporting method. Section A.1 describes the major common concepts that facilitate an integrated ODM/Canvas approach to KA for domain engineering. Section A.2 gives an overview of the ODM domain engineering life cycle, and discusses the specific interfaces between the ODM process and the planning process encompassed by Canvas. Section A.3 presents some further guidelines to consider in carrying out Canvas-style KA planning within the ODM context.

A.1 ODM and Canvas: Common Concepts

ODM reflects the idea that domains are socially defined agreements about an "intended scope of applicability." Domains are always grounded in some "organization context," a community of interest that can take many different forms, such as a company or division, a consortium of multiple organizations or a standards organization. Stakeholder issues, always a potential problem in any project, are critical risk factors in domain engineering, which by definition involves designing for multiple contexts of use. ODM begins with systematic exploration of stakeholder interests to guide selection, scoping and definition of domains strategically aligned with the business interests for the organization(s). This stakeholder context provides further grounding throughout the subsequent phases of the process.

One of the primary purposes of domain modeling according to ODM is *context recovery*, which is the process of identifying and making explicit the assumptions embedded within software artifacts. These assumptions come from the cultural context in which the artifacts are developed and used. This means that domain modeling in ODM bears considerable similarity to cross-cultural investigation of the sort found in ethnographic studies. However, because of the particular constraints of domain modeling for software, knowledge acquisition for ODM imposes some general requirements that go beyond standard practice and general principles and skills of ethnography.

ODM shares with traditional ethnography an attention to the effects of bias on the information gathered from the culture to be studied. On the other hand, the ODM approach recognizes a wider range of stakeholders than many other ethnographic endeavors. Whereas any ethnographer tries to impose little on the culture and to be aware of the politics of "us" and "them," little work has been done in ethnographic research that is equivalent to what the management and consulting world would call alliance forming and management, consultant-client contracting, or stakeholder analysis. Canvas provides support for the traditional ethnographic needs of ODM by recognizing knowledge acquisition as a process of communicating between work cultures. Canvas also provides for more than a simple two-player (us and them) stakeholder model in its consideration of the various stakeholder cultures.

The ODM approach is centered much more on the study of variability than is common in ethnographic studies. Since the result of a domain modeling project according to ODM is an asset base that can be used by many stakeholders in many contexts, the variability inherent in the domain must be elicited, maintained and managed throughout the ODM lifecycle. Canvas specifically provides support for eliciting and representing variability.

In domain engineering, it is important for the domain engineer to see a wide network of stakeholders and also to see where he resides and where the informants reside. He must also apply that knowledge to differing stakeholder roles, interests and domain engineering objectives. This is important for insightful and accurate data analysis and interpretation that captures how the players and their perspectives are situated socially by their context. This ability influences important choices in the planning and modeling phases of ODM. Canvas specifically draws its informants from various stakeholder groups, and different kinds of practitioners in those groups.

ODM views knowledge acquisition as an intervention into the focus community, and includes within its knowledge acquisition guidelines a cycle of elicitation and representation that continually accounts for feedback of information to the informant. The Canvas model of knowledge acquisition as intervention fits well into this pattern, by recognizing the effects that the process of knowledge acquisition itself can have an impact on the focus community.

A.2 Interface between ODM and Canvas

The ODM domain engineering method is structured as a core process model and a set of supporting methods. The core ODM process is structured into three main phases: *Plan Domain*, *Model Domain*, and *Engineer Asset Base*. ODM phases are iteratively decomposed into sub-phases and tasks; each task in turn is documented via sequences of or alternative activities and a structured set of work products that integrate information gathered throughout the entire process. Major results of each phase — the domain definition, domain model, and asset base — are targeted to provide direct benefit to the project stakeholders, as well as to provide a systematic starting point for the next phase.

Information Acquisition techniques are a major supporting method area within ODM, which fit into the overall ODM process within the sub-phase called *Acquire Domain Information*. Information is gathered throughout the process; but it is within this sub-phase where being systematic about data-gathering becomes critical. Exhibit 23 shows the task structure of the ODM process, and where the *Acquire Domain Information* sub-phase is located.

The scope of Canvas is broad enough to allow its use as a supporting method of ODM for the *Plan Data Acquisition* task within the *Acquire Domain Information* sub-phase. Within the core ODM method, the *Plan Data Acquisition* task involves several key activities:

- Selecting a representative set of systems to study to produce the domain model;
- Selecting the settings that will be studied for each representative system;
- Selecting the information sources that will be consulted for each representative system and setting. Consultation of sources could include interviewing human informants, observing domain-related processes, or analyzing artifacts;
- Characterizing biases of investigators.

Canvas offers guidance to support each of these activities, even though it does not provide a detailed process model for the data acquisition planning process. This section will describe how to use Canvas principles as a supporting method for the ODM *Plan Data Acquisition* task.

Knowledge acquisition goals (required as part of the planning process) will be shaped in large part by the fact that the KA enterprise supports a domain engineering project. This means, for example, that management and documentation of variability will be an essential part of the pro-

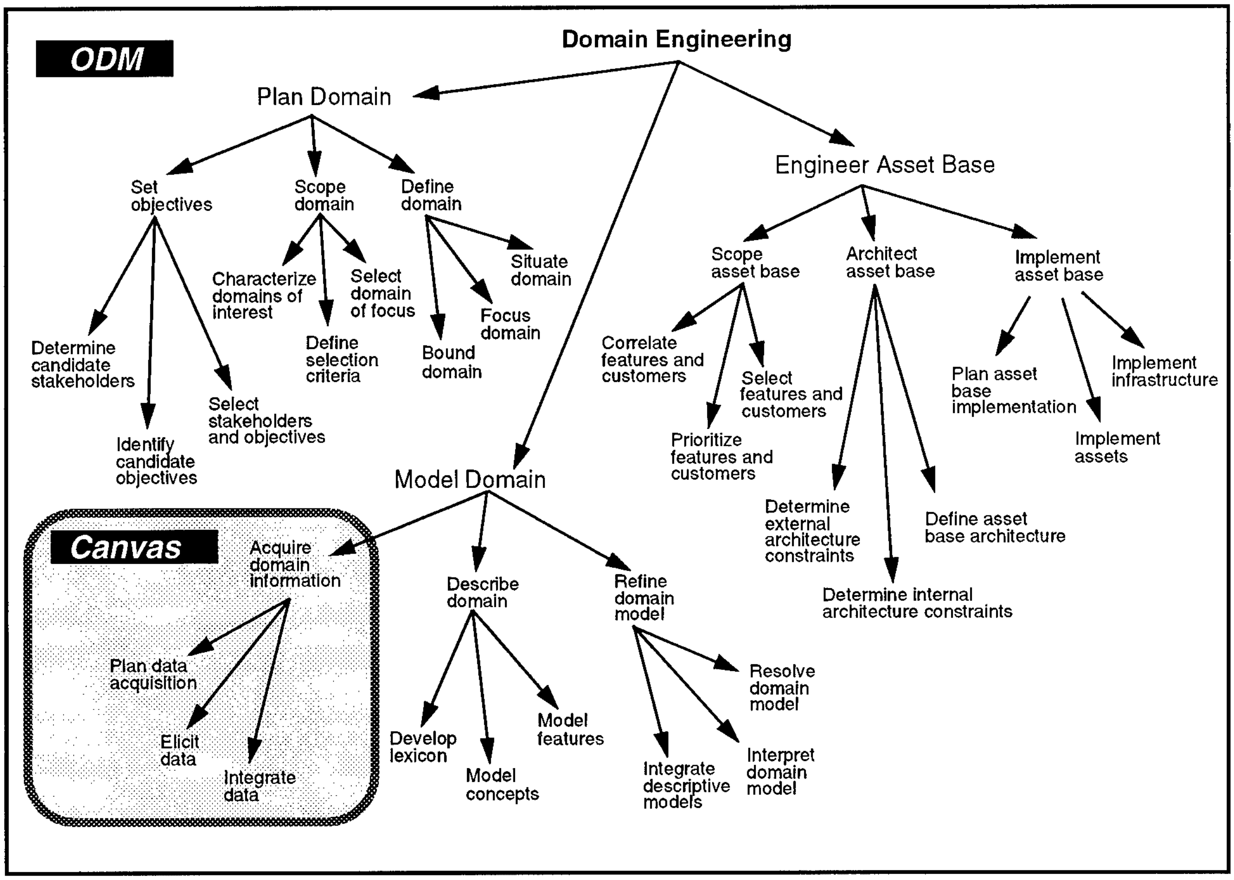

Exhibit 23. Canvas as a Supporting Method of ODM

cess. It also means that, for domain engineering projects in software-intensive domains, KA techniques specifically addressing technology-intensive settings will be required.

When invoked in the ODM context, Canvas KA planning has a number of ODM workproducts and data items available as controls and inputs that can support the planning process. These are described below. A full description of these workproducts and data items, their content and development can be found in the ODM Guidebook V2.0 [46]. In this section, we will follow the conventions of [46] by printing workproducts in SMALL CAPITALS.

ODM Controls to Canvas

The following workproducts and data items constitute controls on the *Plan Data Acquisition* task:

- PROJECT OBJECTIVES. In addition to the general KA goals established due to the domain engineering nature of the project, the specific objectives of the project will have direct impact on the KA enterprise goals. For example, a domain engineering project may be initiated to capture veteran expertise in application development. In this case, the dossier created by Canvas (which will be called the DOMAIN DOSSIER in ODM) will be of direct use to the organization as codified documentation of expertise. If PROJECT OBJECTIVES include the discovery of innovative features that might provide competitive advantage to a software product line, then direct observation of end-users interacting with legacy systems might take place, to elicit opportunities for new features.

- PROJECT CONSTRAINTS. These constraints can include budget, schedule, access constraints as well as constraints such as use of proprietary data.

  Up-front acknowledgment of and planning within constraints is part of any systematic KA process. In general, it is very easy to spend too much time gathering data. Without a budget and plan that shows how much information can actually be used, it is easy to expend scarce resources collecting data that will not be able to be effectively utilized in modeling. Similarly, it is easy to base an initial plan on the assumption that access to some key informants will be forthcoming, only to find that such access is extremely difficult to obtain in reality.

  Constraints are particularly relevant in the domain engineering context. Data is generally being collected from multiple exemplars, which means that the focus of data collection and level of detail become essential controls, since an error in gathering too much data may be repeated on multiple systems. Data may be collected for a narrow domain, which may not be familiar to practitioners as a conventional way of dividing up the functional concerns of the system(s) under investigation.

- DOMAIN DEFINITION. In ODM domain engineering, the DOMAIN DEFINITION provides a tight focus for data acquisition, circumscribing both a specific set of exemplar systems that are included in the domain, and a specific focus area within the overall functionality of systems in the domain. This is one advantage of using Canvas in the ODM context as opposed to within a general system engineering context.

  This does not mean that in a given interview only domain-relevant data will be collected. Informants may not have access to or understanding of the concept of the domain boundary (as refined in the earlier *Define Domain* sub-phase of ODM) as a basis for discourse or focusing attention. Information must be acquired in chunks and sequences that the informants find workable, and subsequently filtered. Yet the DOMAIN DEFINITION can still help investigators to steer inquiry in more focused directions and to rapidly filter the data acquired to extract domain-relevant material, while the session is still comparatively fresh in their experience.

  When ethnographers gather data about a cultural scene, they may be fairly open about what topics will emerge as an area of focus. An anthropologist may be specifically interested in cultural areas such as kinship structures and direct information gathering towards these topics. When using KA as a kind of pre-requirements analysis for system building, there is a danger of lack of focus. To some extent all work practice taking place in the examined settings may appear to be relevant for knowledge acquisition. Though technologists intending to build systems are the primary audience for the KA materials, it may be difficult to know how to filter the materials acquired to give the technologists only the relevant data. There may be an indistinct notion of what technology is intended to be introduced. There may be only weak correlations between technology specifications and the work practices that must be studied in order to assess the viability of the technology. Particularly for innovative technology, it may be very difficult to get end users to provide data about system capabilities which they have not yet used or even seen. The implication is that technology development goals form poor mechanisms for bounding and scoping knowledge acquisition. Investigators are likely to set themselves the goal of describing complete scenarios of work practice within the settings, with the aim of getting the data first and deciding on relevance later.

ODM Inputs to Canvas

The following ODM workproducts and data items serve as inputs to the KA planning processes addressed in Canvas:

- DOMAIN STAKEHOLDER MODEL, STAKEHOLDER DOSSIER

The DOMAIN STAKEHOLDER MODEL and the STAKEHOLDER DOSSIER contain information about the stakeholders for the domain engineering project. The dossier includes all information about the stakeholders and their positions, and can include such items as company prospectus, commercialization plan, interview notes with key personnel, etc. The model includes the relationships among the stakeholders and their roles in the project. This information is not only needed to determine what information will be gathered; it is also the primary source of information for finding informants who will be willing and able to participate in the knowledge acquisition project.

- DOMAIN STAKEHOLDER KNOWLEDGE

  This data item refers to all information that is available from stakeholders. For the purposes of knowledge acquisition, it is the stakeholders who are chosen as informants who are of interest; the domain stakeholder knowledge then refers to informant knowledge. Access to this information is gained in Canvas mainly through interviews.

- EXEMPLAR SYSTEM ARTIFACTS, DOMAIN ARTIFACTS

  The other main source of information in Canvas is artifacts that are examined. ODM divides artifacts into two categories. EXEMPLAR SYSTEM ARTIFACTS are documents, notes, code fragments, executable systems, etc. pertaining to a single software system to be studied. DOMAIN ARTIFACTS are any materials pertaining to the domain, that are not related to any particular system. Examples of such artifacts are survey articles, consumer's reports, or even models created by previous domain engineering efforts. Canvas makes use of both of these types of information, though it tracks their threads differently.

The following Canvas results contribute to ODM workproducts that are derived from the *Acquire domain information* sub-phase:

- KA objectives called out in this document in Section 4.1.1 correspond closely to the "data acquisition goals" created as part of the DATA ACQUISITION PLAN in ODM. These should be aligned closely to the overall PROJECT OBJECTIVES.
- The list of elements selected in Section 4.1.3 constitute the REPRESENTATIVE SYSTEMS SELECTION in ODM. The criteria used in Canvas for selecting these representative systems draws upon information available in the STAKEHOLDER DOSSIER and the DOMAIN STAKEHOLDER MODEL.
- The structure for the dossier that is determine in Section 4.1.5 is the "bootstrap" for the DOMAIN DOSSIER that is built and maintained throughout the *Acquire domain information* sub-phase in ODM.

A.3 Guidelines for Integrating Canvas and ODM

Although Canvas fits very well with a domain engineering project using ODM, Canvas was designed to be more comprehensive in its scope. This means that there are particular activities in Canvas about which we can say more, when we know that Canvas is being used in an ODM context. Typically, since ODM provides more structure to the process than would be present in an undifferentiated setting, the changes are usually to relieve the Canvas team from some worries that are dealt with in other processes in the ODM context. On some occasions, both ODM and Canvas cover similar territory, and the combination allows for a subtle distinction, and a choice of whether a particular decision should be made as part of Canvas or as part of ODM.

Representation versus modeling

In Canvas, the representation of knowledge plays a key role in the process, since it is through representation that knowledge is made available for verification by the focus community, or transferred to the target community. Depending on the choice of representational notation, the process of representing knowledge could be similar to a modeling process. The extent to which modeling must be done depends on the characteristics and requirements of the target community; the target community might require complete, semantically sound models in the dossier.

When Canvas is performed in an ODM setting, we know that the subsequent ODM domain modeling sub-phase creates a model from the knowledge in the dossier. This means that certain modeling activities need not be performed as part of knowledge acquisition, but can be left for the domain modeling stage of ODM

Variability management

Probably the most important example of the previous point is variability. The act of making comparisons and determining just what parts of the knowledge represent commonality and what parts represent variability is a demanding task. Since the Canvas approach stresses that variability be handled as part of the knowledge acquisition effort, when Canvas is performed on its own explicit representation of variability will have to be included in the dossier.

When Canvas is performed as part of an ODM domain modeling project, there is a systematic process in which variability will be modeled. This does not mean that knowledge acquisition practitioners using Canvas in an ODM setting can just forget about variability! The treatment of variability during knowledge acquisition in Canvas includes realizing that the knowledge represented need not conform to some common agreement among diverse sources. When Canvas is performed in an ODM context, the varying viewpoints must still be collected; only the detailed modeling of which information represents differences and which represents commonalities can be delayed until the later ODM *Describe domain* sub-phase.

Target audience is modeling team

When knowledge acquisition is done in an ODM context, the main audience for the dossier is already set, namely, the team who will be doing the domain modeling. This team might be the same team who is doing the knowledge acquisition, or might be another group. In any case, the set of representational notations with which the team is familiar will include some modeling notations, and it can be assumed that the audience is "model savvy".

Two levels of scoping issues

Canvas provides one point at which the sources of information that will be consulted in knowledge acquisition are determined (Section 4.1.3.3 and Section 4.1.3.4). These selections are based primarily on properties of the information source, including issues of currency and accessibility.

In ODM, there is another chance to scope the set of knowledge sources to be studied. As part of the *Plan Domain* phase of ODM, stakeholder issues are used to determine what belongs in the domain and hence, what is a candidate for being selected for study during knowledge acquisition. This means that when performing Canvas as part of an ODM project, there is more guidance for deciding what will be studied and what will not.

Determination of topics of focus

In a similar way, the domain scoping phases of ODM that precedes data acquisition provide a framework for determining topics of focus. Outside an ODM context, some other method must be used to determine what topics should be the focus of elicitation.

Appendix B: Representing the Knowledge Acquisition Process

The production of Canvas from information that was previously only available to the ODM and SEP teams individually involved a process of knowledge acquisition itself. Along the way, we used our own tools to help us to keep track of the material we were collecting. In particular, we used RLF, a hierarchical modeling notation often used in conjunction with ODM, to represent the relationship among the terms in the lexicon and in the definition of knowledge acquisition according to Canvas. We also produced an interaction diagram, similar to the ones common in SEP, to describe how the various elements interact.

B.1 RLF Model of Knowledge Acquisition

This section presents the fundamental concepts used in Canvas, modeled using KNET [10] representations. This section can be used in conjunction with Appendix C: "Canvas Lexicon" to gain a broad understanding of how these concepts are used in Canvas. The hierarchical aspects of KNET have been used to show the inclusion (is-a) relationships of the various terms to one another, while the relationships of KNET have been used to show other relationships. In this appendix, we are using a very small subset of the full capabilities of KNET.

KNET provides the basis for the underlying semantic network representation of the Reuse Library Framework (RLF), and is derived from knowledge representation languages like KL-ONE. Many readers may already be familiar with KL-ONE style languages; however KNET has a few features which we will be using in this appendix that might not be familiar to users of other knowledge representation languages. For readers who are not familiar with these languages, the following description can be used as a primer to understand this appendix. Further information can be found in the RLF documentation [49] and [50].

B.1.1 Fundamentals of KNET

In KNET, the principal entities are categories, objects and relationships. A category models a class of things, such as the class of all learning interactions or the class of all Knowledge Acquisition Sessions. A relationship describes the structure and properties of categories. This appendix does not make use of KNET objects.

Categories in KNET are organized into a specialization hierarchy. A category *A* specializes another category *B* if *A* represents a subset of the class described by *B*. This means that the most general category appears at the top of the hierarchy with more specific categories below it and the most specific categories at the very bottom. In this appendix, specialization will be shown by a straight bold arrow from the specific category to the more general.

Relationships in KNET describe the structure and properties of categories. For instances, a Knowledge Acquisition Session is performed by an Investigator and produces a Workproduct. Such qualities are represented in KNET by associating relationships with a category. For example, Knowledge Acquisition Session includes relationships for knowledge source and performed_by. In this appendix, relationships will be shown as a thin broken arrow from one category to another. The break is annotated with the name of the relationship.

Relationship differentiation allows a relationship to be described in a more detailed way than is possible with a single relationship. Differentiation can thought of as specialization of relation-

ships and is denoted with dashed arrows between the "elbows" in two broken lines. For example, a session has as its knowledge source any number of authorities; the authorities who are focus practitioners are known as informants, while the authorities that are workproducts are known as artifacts. This captures the notion that "artifact" and "informant" are stances that one takes toward workproducts and focus practitioners respectively, and that these stances are just two special cases of the stance of using any authority as a "knowledge source."

B.1.2 Use of the KNET Model of Knowledge Acquisition

We provide a KNET model of knowledge acquisition terms for three different reasons. It is not necessary to understand the subtleties of KNET to obtain all of these benefits.

1) The specialization hierarchies help to clarify relations between terms. The definitions of terminology given in the lexicon (Appendix C) make heavy use of other terms. Many of the relationships between terms are captured in the hierarchy structures of the KNET diagrams.We encourage the reader to flip back and forth from the lexicon to this appendix to see how the terms relate.

2) The relationship networks serve as a summary of the structure of the KA process. Many subtleties, such as the fact that a KA workproduct has a particular audience, which is a community of practice, are explained in the text. Once these explanations have been absorbed, the relationships in the KNET diagrams are a concise way to record and present this information.

3) The KNET model can be used as a "starter set" for a dossier of knowledge acquisition workproducts. Any aspect of the planning process could be used to index the dossier. The most likely parts of the KNET model to be used as an index are the categorizations of audiences and the types of representation.

B.1.3 KNET Models of Knowledge Acquisition

The KNET diagram for the entire knowledge acquisition process is too complex to display in a single display; we have therefore broken it into three pieces. Each piece is focused on a single category. All specializations of that category are shown, as well as all relations defined for the focus category or any of its specializations (i.e., any relation arrows that lead *away* from the category of focus or its specializations.) Any category that must be shown in order to display these relations is also shown. There is no guarantee that specialization hierarchies for non-focus categories are complete as shown. In order to find complete specializations for a non-focus category, it is necessary to look in the diagram where that category, or one of its generalizations, is the focus.

Exhibit 24 shows the KNET diagram for the knowledge acquisition session itself. A knowledge acquisition session is performed by an investigator, has a knowledge source that is an authority (about some topic, which is not shown), and produces a workproduct. In the case where the knowledge source is a human being, the session is called an interaction. When the knowledge source is a workproduct, the session is a study. Artifacts and informants are differentiated forms of knowledge sources. When describing a session, we typically refer to informants, investigators and knowledge sources, rather than focus practitioners, workproducts and authorities because we are interested in their roles in the session.

Exhibit 24 shows the KNET diagram for the practitioners involved with the knowledge acquisition effort. A practitioner is a member of some community of practice. The two specializations of practitioner, focus practitioners and investigators, are members of the focus community and investigator community respectively. When the meaning is clear from the context, we often refer to focus practitioners simply as practitioners.

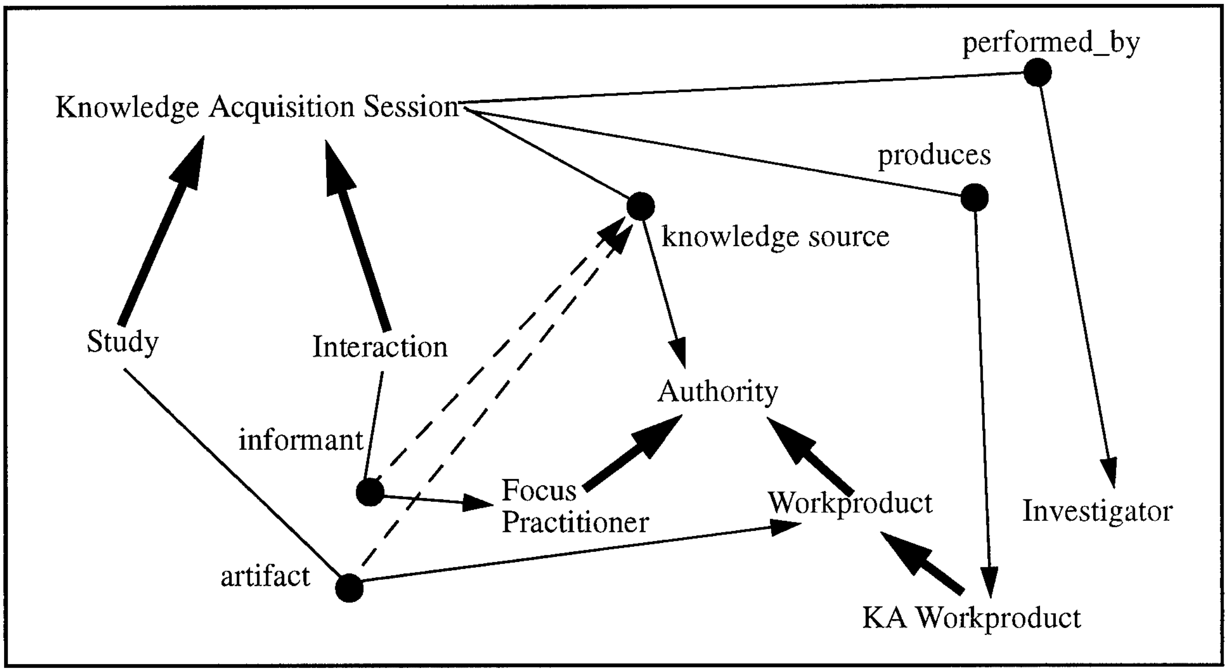

Exhibit 24. Relationships Centered around the KA Session

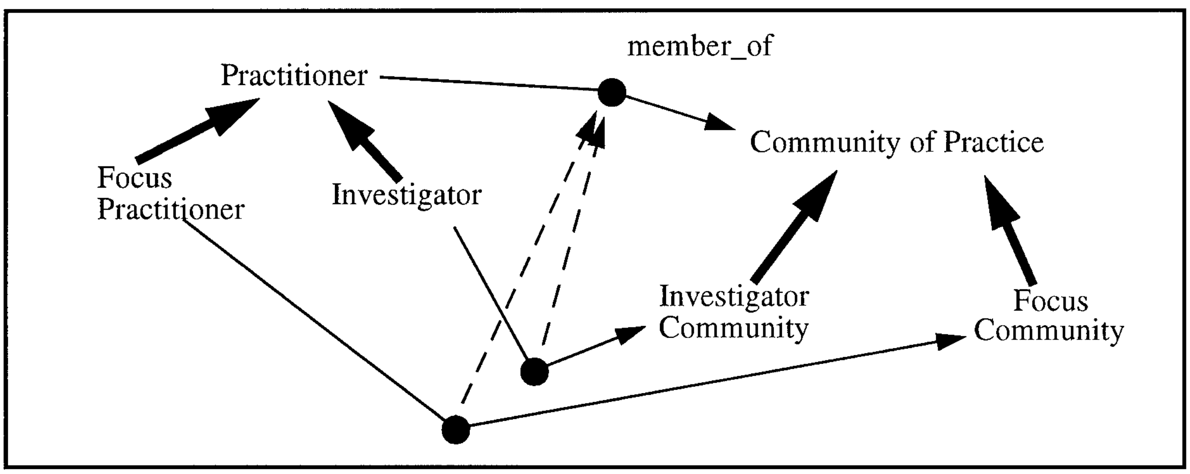

Exhibit 25. Relationships Centered around Practitioners

Exhibit 24 shows the KNET diagram for the community of practice. In Canvas, we are interested in three kinds of communities: an investigator community, a focus community, and a target community. Every workproduct has an audience community, and the members of each community practice in some work setting.

B.1.4 Epilogue: Using KNET as a Knowledge Acquisition Tool

KNET is an example of a taxonomic notation, as described in Section 5.2.4. In this appendix, we have used a subset of KNET (i.e., without cardinality or restriction of cardinality, and leaving out different types of differentiation) as a display mechanism. In order to make the diagrams legible, we have added some extra parameters, as described in Section B.1.3, to control the display. The

Exhibit 26. Relationships Centered around Community of Practice

result is a notation that could be used for a knowledge acquisition workproduct. Here, the topic of the workproduct is the knowledge acquisition process itself. Another representation of much the same material can be found in Appendix C in another notation, that of a lexicon.

The audience of this workproduct is you, the reader of this guidebook. However, there might be other interested parties who will become an audience for the dossier. In this case, the authors of this guidebook were an audience for the diagrams. As is often the case, we began examining the diagrams as a way of validating the information we were gathering. As time went on, however, we began to use them as a reference for our own work while writing the guidebook.

Each notation brings with it certain views on the information that others do not. For example, in Exhibit 24, we needed to find a name for the category of entities that could serve as knowledge sources. The two known specializations of this new category, namely focus practitioner and workproduct, had stories to be told about them in their own rights (namely, as the subject of two of the threads in the Canvas.) However, the new category had not played a role in the Canvas exposition so far (and did not appear in the lexicon). The formal constraints on the KNET diagrams forced us find a name. This process revealed any number of stakeholder issues that had not been evident before (such as why "knowledge source" had to be a relation, not a category; which category the name practitioner should refer to, etc.) These issues, in turn, were fed back into the rest of the writing of this guidebook.

B.2 Interaction Model of a Knowledge Acquisition Session

Canvas is a melding of SEP and ODM. A top level view of the interactions of Canvas are shown in the interaction diagram of Exhibit 27. An interaction diagram is a variation on the petri net. The primary distinction between interaction diagrams and classical petri nets is that objects (items represented in ovals) may retain their identify through transitions (rectangles). Transitions are places where objects are created, destroyed, modified, change state, or interact with one another. The iterative nature of the interactions is evident.

Sessions conducted by the investigator(s) produce work products using as input artifacts from the domain and/or the information provided by interviews and/or observations involving informants from the domain. Some artifacts are carried over from prior work in ODM (i.e., legacy systems, design specifications, user manuals, and other documents.) Other artifacts are derived from prior work in SEP, such as representative scenarios, concepts, and task decompositions.

Transfer is the process that models the incorporation of Canvas work products into the understanding of the domain by the audience.

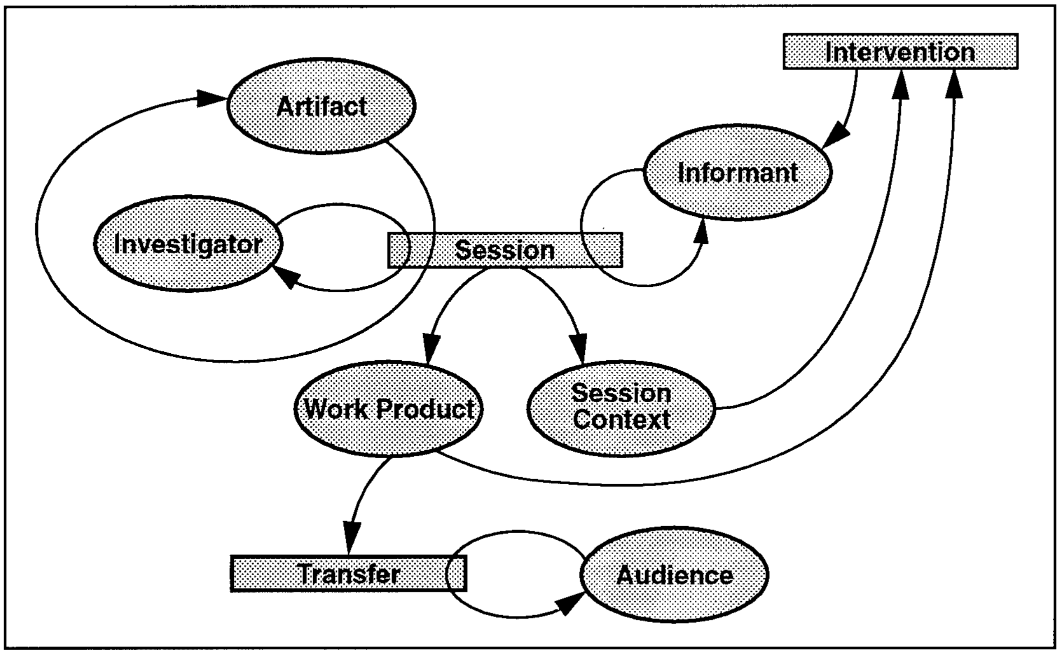

Exhibit 27. Interaction Diagram for a Knowledge Acquisition Session

The completion of the feedback loop for investigators provides for iterative improvement of the domain context held as corporate knowledge by the investigator community. The production of the KA workproduct, and the context of the setting itself, can intervene on the informant's work practice, either individually, by clarifying his own understanding of his field, or as a community, by fostering connections that had not otherwise been made.

SEP and ODM each embody an iterative improvement approach to domain analysis, with variations in emphasis. SEP tends to concentrate upon the description of the performer-specific part of domain analysis. In contrast ODM tends to focus upon legacy and emerging system analysis for the purpose of establishing systematic software reuse strategies. Their combination provides a balanced approach to the analysis of the entire enterprise.

Appendix C: Canvas Lexicon

Lexicon Conventions

This lexicon defines terms used in the document. Within the main body of the document, these terms when first referenced in any given section appear in ***bold italic*** type. In later references within the same section they will generally appear in regular type, but may occasionally be repeated in ***bold italic*** type.

Three types of terms are included in the lexicon:

- *General descriptive terms* for concepts essential to Canvas. We have tried to include only terms that are used with special meaning in the Canvas context. Some of these terms are used frequently in other disciplinary contexts (e.g., "informant" has typical conventions of usage in social scientific research) or general usage (e.g., "audience"). These terms are used in a specific technical sense in Canvas, i.e., to serve as an umbrella for several more specific terms, or to make key distinctions that would otherwise be difficult to convey.

  If you see a familiar term in ***bold italic***, make sure you are clear about the specific meaning it has in the Canvas context. For example, the term "interaction" is used here as a technical term for a person-to-person encounter between an investigator and an informant, in contrast to a study which involves and investigator and an inanimate source of information like a document.

- *Formal and informal terms*. In many cases a term like "session" is used extensively throughout the document, when in fact the meaning of the term is "knowledge acquisition session." In these cases, we have chosen to provide the main definition entry for the *more* formal term. Brackets are used in the more formal term to indicate variant phrases that might be used for the defined term, or in some cases, the fuller and more formal phrase elided in the common usage within the document. For example, in the entry "[knowledge acquisition] session", "knowledge acquisition" is bracketed to show that the term sometimes appears in the text as just "session". A cross reference is included in the lexicon under the less informal term (e.g., "session").

- *Non-Canvas terms*. A few entries may be included in the lexicon for terms used in general reuse literature that do not have special meaning in the Canvas context. These are included for clarity, context and completeness, not to suggest formal definitions for these terms.

In addition to the textual definitions of terms provided in this appendix, Appendix B includes formal notations of the semantic relations between some of the key terms in the lexicon. We encourage the reader to flip back and forth between these two appendices to see how the terms relate.

Lexicon Terms and Definitions

activity

A process or work task performed by a practitioner within a work setting. Activities can be observed by investigators as one form of knowledge acquisition. Note that within the SEP methodology this term has a more constrained and formal definition as part of scenarios. Within Canvas the main significance is that the activity serves as the effective knowledge source for sessions where there is no explicit interviewing of an informant or study of an existing artifact. See also practitioner, work setting, observation.

artifact

A workproduct developed within a work setting, that is used to provide information about the work setting in which it was created or used. An artifact is a type of knowledge source. See also workproduct, work setting, knowledge source.

audience

For any particular knowledge acquisition workproduct, its audience is anyone who can examine the workproduct, and thereby gain some of the knowledge that was the subject of the learning process that produced the workproduct. The audience can be drawn from any community. In some cases the audience community is the same as the knowledge source community or the investigator community. See also knowledge acquisition workproduct and learning.

Canvas

A knowledge acquisition approach that encompasses the idea that each participant (informant, investigator) in knowledge acquisition will be influenced by the ongoing knowledge acquisition effort. The name "Canvas" come from the image of weaving together "threads" corresponding to the lifecycle of each participant. Each thread is monitored and managed for the effect of changing biases throughout the lifecycle. In Canvas, the knowledge acquisition effort is viewed as a way to bridge the cultural gap between two (or more) communities of practice, by explicitly handling the variance as well as the commonality of view both within a single community and between communities. See also participant, knowledge acquisition, thread, knowledge acquisition effort, community of practice.

codification

The transfer of knowledge into a workproduct. Knowledge has been codified into a workproduct, if that knowledge can be learned through examination of the workproduct by a practitioner in the intended audience for the workproduct. See also workproduct.

collaboration

A particular type of learning process resulting from a session with an informant in which the investigator gains new knowledge that was not already held by the informant. Typically in such a situation, the informant learns also. In contrast to reflection, the new knowledge contains some aspects of information unavailable to each participant; hence it could not have been produced by any single participant alone. See also learning, informant, investigator, reflection, transfer.

community of practice

A group of practitioners that share terminology, knowledge, and a set of behaviors and interests in a certain domain. See also work setting.

derivative workproduct

A workproduct that was produced by interpreting some other workproduct, including another derivative workproduct. See workproduct, knowledge acquisition session, artifact.

dossier

A collection of all the knowledge acquisition workproducts created as part of a knowledge acquisition enterprise. Includes links to the artifacts and informants used as knowledge sources for the enterprise. See also knowledge acquisition workproduct, knowledge acquisition enterprise.

embedded [knowledge]

Knowledge that is not held in a form that can be easily articulated by any single member of a community of practice. The knowledge may exist in close alignment with social interactions. For example, an operating room team may have certain competencies in common that were developed over long periods of intensive work practice. The knowledge may be informal and not codified, passed on by direct contact and experiential training. The knowledge may also be embedded because it is part of the surrounding cultural context and hence not deemed as a characteristic aspect of the work setting itself, but is of significance to the investigator and/or target communities.

enterprise

See knowledge acquisition enterprise.

focus community

The community of practice in which the focus of interest is embedded. Practitioners from this community are selected as informants. Contrast to investigator community, target community. See also focus of interest, informant.

focus of interest

The topic in the knowledge source's work setting about which the investigator wishes to acquire information. The knowledge source should be knowledgable about this topic. See topic, knowledge source, work setting.

informant

A practitioner who provides information or knowledge that is embedded in his corresponding work practice setting, to investigators during a knowledge acquisition effort. Knowledge acquisition specialists in the knowledge engineering field tend to use the term "domain expert" to describe the people that they interview in knowledge acquisition. We prefer a less specific term, since some people interviewed will neither consider themselves experts nor be considered such by their community. See also knowledge source, practitioner.

interaction

A session where two people exchange information. Contrast with study, where people interact with artifacts. See also session.

investigator

A member of a community of practice who performs sessions to acquire information that is currently embedded in some *other* work setting. See also session, knowledge source, audience.

investigator community

The community of practice that is performing the knowledge acquisition. Contrast with focus community and target community. See also knowledge acquisition session.

KA

See knowledge acquisition.

knowledge acquisition

Knowledge acquisition is a special kind of knowledge creation process with the following properties:

- Knowledge is elicited from a knowledge source from one work setting by an investigator from another work setting.
- Some portion of the elicited knowledge is codified into a knowledge acquisition workproduct.

The knowledge elicitation and codification activities take place in a knowledge acquisition session. They could take place in a single or in separate sessions.

The knowledge that has been effectively codified is that knowledge that can be learned through examination of the resulting knowledge acquisition workproduct, by a person who was not present at the knowledge acquisition session, but who is in the intended audience for the workproduct. In any knowledge acquisition session, there is always more knowledge created than is effectively codified in the resulting workproduct. See also knowledge source, investigator, work setting, knowledge acquisition session, knowledge creation, codification.

knowledge acquisition effort

That portion of an overall project concerned with the systematic acquisition of knowledge. The effort might be part of an expert systems development project, a conventional systems development effort utilizing knowledge acquisition techniques, a domain engineering effort for the purposes of engineering reusable software assets, or for non-technical purposes such as the gathering of social scientific data. The effort encompasses the people (e.g., investigators, informants, sponsors, etc.), the processes (e.g., interviews, studies of artifacts), and the resulting knowledge acquisition workproducts.

[knowledge acquisition] enterprise

The aspects of a knowledge acquisition effort that are relatively constant throughout the effort (e.g., the objectives, intended audience, dossier infrastructure and resource pools.) Knowledge acquisition planning at the enterprise level is the first and foremost place where stakeholder issues are examined. Contrast to threads and sessions, many of which can be planned within the scope of a single knowledge acquisition effort. See also knowledge acquisition effort, stakeholder.

[knowledge acquisition] objectives

Goals for a knowledge acquisition activity. Objectives can be set for the knowledge acquisition enterprise as well as the knowledge acquisition session. The objectives determined by participants in planning a knowledge acquisition session are primarily those objectives determined by the investigators in the context of the objectives of the knowledge acquisition enterprise of which the session is a part. Objectives for a session include the topics of focus for the session, the knowledge types to be acquired, desired level of detail, etc. Objectives for an enterprise include supplying information to a particular audience. See also topic of focus, knowledge type, knowledge acquisition session, knowledge acquisition enterprise.

[knowledge acquisition] session

A knowledge acquisition session is an event that involves at least one investigator and at least one knowledge source, and generally produces a knowledge acquisition workproduct. For the purposes of Canvas, we only consider sessions where the knowledge source and investigators are associated with different work settings; events within a single setting are the ordinary practice in

that work setting, and are not treated here. The fundamental types of sessions are studies of artifacts and interactions with informants. The audience for the workproduct created may be within the informant's setting, the investigator's setting or a separate target setting. See also investigator, knowledge acquisition workproduct, artifact, informant, work setting, audience.

knowledge acquisition workproduct

A particular workproduct that was created in the work setting of knowledge acquisition, that is, the result of a knowledge acquisition session. The knowledge acquisition workproducts are retained in the dossier. Knowledge acquisition workproducts also have an audience, which is a community of practice who can be expected to be able to comprehend it. See also workproduct, knowledge acquisition session, dossier, audience.

knowledge creation

The most general term for a process that results in new knowledge being created. Individual learning, formal teaching, documenting of process, research, and knowledge acquisition (the focus of this document) are all forms of knowledge creation. See knowledge acquisition.

knowledge source

Anything that can provide information that is embedded in some work settings. Human knowledge sources are called informants, while inanimate knowledge sources are called artifacts. See informant, artifact.

knowledge type

In Canvas, knowledge is typed based on how easily it can be elicited from an informant and how the knowledge is represented.

learning

Any process by which a person gains some knowledge. See also transfer, collaboration and reflection, which are types of learning.

notation

The mode of representation of the information in a workproduct. In general, the notation could be any of natural language, diagrams, charts, tables, mathematical formula or the like. In Canvas, we usually consider workproducts within representations particular to knowledge acquisition. The notation in which a workproduct is represented constrains its possible content, and influences who its audience can be. For example, programming language representations are not usually accessible to medical personnel. See also representation, workproduct.

objectives

See knowledge acquisition objectives.

observation

An investigator eliciting knowledge by being directly present in a work setting and passively observing work activities being performed by practitioners. Whether the observed practitioners are informants or not is a grey area that depends on their degree of awareness of the presence of the investigator, the extent to which their activities are altered as a result of the observation, etc. Also, if a passive recording medium such as audio or video tape is used, one must consider the session in which this knowledge acquisition workproduct is reviewed by investigators as part of

the chain of interpretation for the event recorded. See also informant, knowledge acquisition workproduct.

practice

An activity that takes place within a single work setting. The activity should be considered by the members of the community of practice to be part of the work that goes on in that setting. See also work setting, community of practice.

practitioner

A member of a community of practice. Practitioners are always humans. See also community of practice, informant, artifact.

reflection

A particular type of learning in which the learner gains knowledge without resorting to a knowledge source. Reflection is, therefore, not dependent on a knowledge acquisition session. See also learning, transfer, collaboration, knowledge acquisition session.

representation

A description of some topic, produced during a knowledge acquisition session, and recorded in a workproduct. A representation is written using some notation. See also workproduct, notation.

session

See knowledge acquisition session.

setting

See work setting.

stakeholder

A person, group, or organization that has interests and objectives relative to the knowledge acquisition effort. Stakeholders have diverse viewpoints, experience, and terminology. See also knowledge acquisition effort.

study

An event where an investigator examines a workproduct to learn about a work setting. Contrasted with interaction. See investigator, workproduct, work setting.

systematic knowledge acquisition

A systematic approach to knowledge acquisition provides a repeatable procedure for making key decisions in planning and performing knowledge acquisition, and for recording the results of knowledge acquisition activities in such a way that essential contextual information about the data acquired is preserved.

target community

In a knowledge acquisition project involving three distinct communities, the community that is the intended audience of the generated workproducts. A project in which the audience is either the focus or investigator community will not have a separate target community. When speaking gen-

erally, we refer to the audience community as the target community, whether it is the same as the focus or investigator community, or a third, separate community. Contrast to focus community and investigator community. See also audience, workproduct.

thread

The lifecycle of an informant, artifact, or investigator. The word "thread" is intended to convey the image of linearity of a lifecycle, interacting in a two-dimensional "canvas". Each entity in the knowledge acquisition project (informant, artifact, or investigator) follows a thread through the fabric of the overall project. See also informant, artifact, investigator.

topic [of focus]

A specific criterion for selecting and focusing information gathered in a knowledge acquisition session; part of the knowledge acquisition objectives for the session. See also knowledge acquisition objectives.

transfer

A particular type of learning process resulting from a session with a knowledge source, in which the learner gains knowledge that was already held by knowledge source. See also collaboration, learning.

work practice setting

See work setting.

workproduct

The result of an activity in any work setting. In order to be a workproduct, the result must be in a form that can be accessed by someone who was not involved in the activity. Thus reports, tapes, documents, programs and diagrams are all workproducts, but new ideas that have not been captured on paper are not workproducts. See also work setting, knowledge acquisition workproduct.

[work] setting

An environment where people interact with each other and perform processes. Settings imply a certain stability in that the same people work together on a routine basis. This set of people forms a community of practice. See also community of practice.

Bibliography

In the following bibliography, documents which pertain most directly to ODM and its application are denoted with an asterisk (*) Documents which pertain to SEP and its applications are denoted with a double asterisk (**).

[1] * Army STARS Demonstration Project. Domain Engineering Guidebook. U.S. Army CECOM Software Engineering Directorate, Ft. Monmouth NJ, 1995.

[2] Bailey, K. Typologies and Taxonomies: An Introduction to Classification Techniques. Sage University Paper, Series: Quantitative Applications in the Social Sciences, Series/Number 07-102, Sage Publications, 1994.

[3] Block, P. *Flawless Consulting: A Guide to Getting Your Expertise Used.* Pfeiffer & Co., San Diego CA, 1981.

[4] Boehm, B., W. Scherlis. "Megaprogramming." In *Proceedings of the DARPA Software Technology Conference*, Arlington VA, April 1992.

[5] Bojie, D. "The Storytelling Organization: A Study of Story Performance in a Office-supply Firm." *Administrative Science Quarterly,* Vol. 36, 1991, pp.106-26.

[6] Brown, J. S. "Research that Reinvents the Corporation." *Harvard Business Review,* March-April 1991.

[7] * Collins, P. "Toward a Reusable Domain Analysis." In *Proceedings of the 4th Annual Workshop on Software Reuse*, Herndon VA, November 1991.

[8] Creps, R., M.J. Davis, M. Simos, et al. "Using a Conceptual Framework for Reuse Processes as a Basis for Reuse Planning." In *Proceedings of the 7th Annual Software Technology Conference*, Salt Lake City UT, April 1995.

[9] Foreman, J. "STARS Mission." In *Proceedings of the DARPA Software Technology Conference*, Arlington VA, April 1992.

[10] M. W. Freeman, L. Hirschman, D. P. McKay, F. L. Miller, and D. P. Sidhu. Logic Programming Applied to Knowledge-Based Systems, Modelling, and Simulation. In *Proceedings of the Conference on Artificial Intelligence*, pages 177-193, April 1983.

[11] Frey, D. "Something's Got to Give." New York Times Magazine, March 24, 1996.

[12] **Haddock, G., and K. Harbison. "From Scenarios to Domain Models; Processes and Representations," in *Proceedings of Knowledge-based Artificial Intelligence Systems in Aerospace and Industry. SPIE: International Society for Optical Engineering*, April 1994.

[13] Henninger, S. "Accelerating the Successful Reuse of Problem Solving Knowledge Through the Domain Lifecycle." In *Proceedings of the Fourth International Conference on Software Reuse*, IEEE Computer Society Press.

[14] Johnson, J. *Selecting Ethnographic Informants.* Sage University Paper, Series: Qualitative Research Methods, Volume 22, Sage Publications, 1990.

[15] Jordan, B. "Cosmopolitical Obstetrics: Some Insights from the Training of Traditional Midwives," *Soc. Sci. Med.* Vol. 28, No. 9, pp. 925-944, 1989.

[16] * Kelly, T.P., W. Lam, B.R. Whittle. "Diary of a domain analyst: a domain analysis case-study from avionics." In *Proceedings of IFIP Working Groups 8.1/13.2 Conference on Domain Knowledge for Interactive System Design*, Geneva, May 1996.

[17] Keough, D.M. "Creating the Learning Organization: An Interview with Peter Senge," *The McKinsey Quarterly*, 1992.

[18] Klingler, C. "A Practical Approach to Process Definition." In *Proceedings of the 7th Annual Software Technology Conference*, Salt Lake City UT, April 1995.

[19] Lievegoed, B. *The Developing Organization.* Celestial Arts, Millbrae CA, 1980.

[20] Linde, C. "Reflections on Workplace Learning." Working Paper, Institute for Research on Learning, 2550 Hanover St., Palo Alto CA, 1993.

[21] Mancuso, J.C., and M.L.C. Shaw. *Cognition and Personal Structure: Computer Access and Analysis.* Praeger Press, New York NY, 1988.

[22] Marca, D., and C. McGowan. *SADT, Structured Analysis and Design Technique.* McGraw-Hill, New York NY, 1988.

[23] ** Mettala, Erik, K. Harbison and S. Hufnagel. Scenario-based Engineering Process for Reconnaissance, Surveillance and Target Acquisition, in Proceedings of the ARPA Image Understanding Workshop, Morgan-Kaufman, November 1994.

[24] Moore, G. *Crossing the Chasm: Marketing and Selling Technology Products to Mainstream Customers.* Harper Business, New York NY, 1991.

[25] Noblit, G., R. Hare. *Meta-Ethnography: Synthesizing Qualitative Studies.* Sage University Paper, Series: Qualitative Research Methods, Volume 11, Sage Publications, 1988.

[26] O'Connor. B. *Healing Traditions: Vernacular Health Traditions and their Implications for Clinicians.* University of Pennsylvania Press, Philadelphia, 1994.

[27] Schein, E. *The Clinical Perspective in Fieldwork.* Qualitative Research Methods Series 5, Sage, Newbury Park CA, 1987.

[28] Schein, E. *Process Consultation, Vol. 2: Lessons for Managers and Consultants.* Addison-Wesley, Reading MA, 1987.

[29] Schon, D. *The Reflective Practitioner: How Professionals Think in Action.* Basic Books, new York NY, 1983.

[30] Schwartz, P. *The Art of the Long View, Planning for the Future in an Uncertain World.* Doubleday/Currency, New York NY, 1991.

[31] Schwartzman, H. *Ethnography in Organizations.* Qualitative Research Methods Series 27, Sage, Newbury Park CA, 1993.

[32] Senge, P. *The Fifth Discipline.* Doubleday/Currency, New York NY, 1990.

[33] Senge, P., et al. *The Fifth Discipline Fieldbook: Strategies and Tools for Building a Learning Organization.* Doubleday/Currency, New York NY, 1994.

[34] Shaw, M. L. G. and McKnight, C. *Think Again: Personal Decision Making and Problem Solving*. Prentice Hall, Hemel Hempstead, 1981.

[35] * Simos, M. "The Growing of an Organon: A Hybrid Knowledge-Based Technology and Methodology for Software Reuse." In *Domain Analysis and Software Systems Modeling*, R. Prieto-Diaz and G. Arango, ed., IEEE Computer Society Press, 1991.

[36] * Simos, M., "Juggling in Free Fall: Uncertainty Management Aspects of Domain Analysis Methods." In *Proceedings of the 5th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Springer-Verlag, July 1994.

[37] * Simos, M., "Organization Domain Modeling (ODM): Formalizing the Core Domain Modeling Life Cycle." *SIGSOFT Software Engineering Notes*, Special Issue on the 1995 Symposium on Software Reusability, Aug 1995.

[38] * Simos, M., "Domain Modeling Techniques for Representing Commonality and Variability: Towards a Comparative Framework." *Proceedings of the 7th Annual Workshop on Software Reuse*, St,. Charles, IL, August 1995.

[39] Softech, Inc. "Integrated Computer-Aided Manufacturing (ICAM) Architecture Part II. Volume IV - Function Modeling Manual ($IDEF_0$)." Technical Report AFWAL-TR-81-4023 Volume IV, Materials Laboratory (AFWAL/MLTC), AF Wright Aeronautical Laboratories (AFSC), Wright-Paterson AFB, Dayton OH, June 1981.

[40] Spradley, J. *The Ethnographic Interview.* Holt, Rinehart, and Winston, New York NY, 1979.

[41] Spradley, J. *Ethnographic Observation.* Holt, Rinehart, and Winston, New York NY, 1979.

[42] * STARS. Army STARS Demonstration Project Experience Report. STARS Technical Report STARS-VC-A011R/003/02, STARS Technology Center, Arlington VA, April 1996.

[43] STARS. STARS Conceptual Framework for Reuse Processes (CFRP), Vol. I: Definition, Version 3.0. Unisys STARS Technical Report STARS-VC-A018/001/00, STARS Technology Center, Arlington VA, October 1993.

[44] STARS. STARS Conceptual Framework for Reuse Processes (CFRP), Vol. II: Application, Version 1.0. Unisys STARS Technical Report STARS-VC-A018/002/00, STARS Technology Center, Arlington VA, September 1993.

[45] STARS. Learning and Inquiry Based Reuse Adoption (LIBRA): A Field Guide to Reuse Adoption through Organizational Learning, Version 1.1. Loral Defense Systems-East STARS Technical Report STARS-PA33-AG01/001/02, STARS Technology Center, Arlington VA, February 1996.

[46] * STARS. Organization Domain Modeling (ODM) Guidebook, Version 2.0. STARS Technical Report STARS-VC-A025/001/00, Lockheed Martin Tactical Defense Systems, Manassas VA, June 1996.

[47] STARS. Reuse Strategy Model: Planning Aid for Reuse-Based Projects. Boeing STARS Technical Report D613-55159, STARS Technology Center, Arlington VA, July 1993

[48] Trist, E. "Referent Organizations and the Development of Interorganizational Domains." *Human Relations*, Vol. 36, No. 13, pp. 269-84, 1983.

[49] * Unisys. Reusability Library Framework AdaKNET and AdaTAU Design Report. PAO D4705-CV-880601-1, Unisys Defense Systems, System Development Group, Paoli PA, 1988.

[50] Unisys. RLF Modeler's Manual, version 4.2, Unisys Defense Systems, System Development Group, 1993.

[51] ** Wang, Wei, Steve Hufnagel, Pei Hsia, Seung Min Yang, "Scenario Driven Requirements Analysis Method." *Proceedings of the Second Internations Conference on Systems Integration*, June 15-18 1992, Morristown, NJ, IEEE.

[52] Weisbord, M. *Discovering Common Ground*. Berrett-Koehler, San Francisco CA, 1992.

[53] Weizenbaum, J. *Computer Power and Human Reason*, W. H. Freeman and Co., San Francisco, 1976.

[54] Welbank, M. British Telecom Report on Knowledge Acquisition, British Telecom, London 1983.

[55] Winograd, T., F. Flores. *Understanding Computers and Cognition: A New Foundation for Design*. Addison-Wesley, Reading MA, 1987.

[56] Zuboff, S. *In the Age of the Smart Machine*. Basic Books, New York NY, 1988.